3-24-2016

Using Artificial Neural Networks to Predict Disease Associations for Chemicals Present in Burn Pit Emissions

Amanda R. Taylor

USING ARTIFICIAL NEURAL NETWORKS TO PREDICT DISEASE ASSOCIATIONS FOR CHEMICALS PRESENT IN BURN PIT EMISSIONS

THESIS

Amanda R. Taylor, MSgt, USAF

AFIT-ENV-MS-16-M-189

DEPARTMENT OF THE AIR FORCE
AIR UNIVERSITY

AIR FORCE INSTITUTE OF TECHNOLOGY

Wright-Patterson Air Force Base, Ohio

AFIT-ENV-MS-16-M-189

USING ARTIFICIAL NEURAL NETWORKS TO PREDICT DISEASE ASSOCIATIONS FOR CHEMICALS PRESENT IN BURN PIT EMISSIONS

THESIS

Presented to the Faculty

Department of System Engineering and Management

Graduate School of Engineering and Management

Air Force Institute of Technology

Air University

Air Education and Training Command

In Partial Fulfillment of the Requirements for the

Degree of Master of Science in Industrial Hygiene

Amanda R. Taylor

Master Sergeant, USAF

February 2016

DISTRIBUTION STATEMENT A.
APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.

AFIT-ENV-MS-16-M-189

USING ARTIFICIAL NEURAL NETWORKS TO PREDICT DISEASE ASSOCIATIONS FOR CHEMICALS PRESENT IN BURN PIT EMISSIONS

Amanda R. Taylor
Master Sergeant, USAF

Committee Membership:

Dr. Willie Harper
Chair

Lt Col Robert Eninger, PhD
Member

Dr. David Mattie
Member

Abstract

In June of 2015, 27,378 of the 28,000 returning Operation Iraqi Freedom/Operation Enduring Freedom (OIF/OEF) veterans report being exposed to burn pits. According to Barth et al. (2014), 9,660 returning OIF/OEF veterans were diagnosed with respiratory diseases, to include asthma, bronchitis, and sinusitis, thus strengthening the need to develop decision support tools that can be used to understand the relationships between chemical exposure and disease. In this study an Artificial Neural Network (ANN) was used to predict the chemical-disease associations for burn pit constituents. Ten burn pit constituents were tested using varying hidden layers, similar chemical structure relationships, and three Training, Validation, and Testing (TVT) ratios. The ANN predicted misidentification rates of 73% or greater when the hidden layer size varied between 1 and 5. Misidentification rates of 75% or greater were observed for ANN simulations when the TVT ratios ranged from 60/20/20 to 80/10/10. ANN-based screening of chemical groups containing chemicals with benzene rings and chemicals containing hydrocarbon chains produced misidentification rates of 73% or greater, and R^2 values of 0.0762 and lower. Hidden Layer size, TVT ratios, and chemical structure had little effect on the model's performance; additional training data is needed to improve the predictive capability of the ANN. The ANN-based screening of individual burn pit constituents produced several chemicals with R^2 values greater than 0.8. These chemicals have been prioritized to further develop predictive ANN models for human health force support, resulting in the first research screening burn pit constituents with an ANN, and the first to prioritize burn pit emissions for future testing.

To Casper and Judy

Acknowledgments

I would like to thank my advisor, Dr. Willie Harper, for his guidance and direction throughout the thesis process. I would also like to thank my advisors Lt Col (Retired) Dirk Yamamoto and Lt Col Robert Eninger for their encouragement and support.

Amanda R. Taylor

Table of Contents

Page

List of Figures

Page

List of Tables

USING ARTIFICIAL NEURAL NETWORKS TO PREDICT DISEASE ASSOCIATIONS FOR CHEMICALS PRESENT IN BURN PIT EMISSIONS

I. Introduction

1. General Issue

The risks associated with industrial use chemicals are reduced as a result of the performance of risk assessments. Establishing risk associated with a chemical involves the identification of the hazard, a hazard assessment, an exposure assessment, and basic characterization of the risk, resulting in the generation of effects caused by chemical exposure. According to the American Chemical Society (ACS), chemical risk assessments cannot be accomplished using a single set of analytical tests. Instead, a chemical risk assessment should involve a process for selecting the most appropriate method to evaluate the impacts associated with the life cycle exposure to a chemical (ACS, 2015).

Due to the evolving nature of the military-related missions overseas, risk assessments are often performed post-hoc. The highly publicized post-hoc burn pit assessments have sparked a need for military chemical risk assessment reform. As a result of the risk assessment reform, this thesis will explore the diseases that are linked to the constituents present in burn pit emissions. The research involved in this thesis focused on studying the relationship between chemicals and disease, burn pit emission constituents, and the use of an Artificial Neural Network (ANN) to predict chemical-disease associations among burn pit constituents.

a. The relationship between chemicals and disease

i. Etiology

The etiology of disease has many facets. One facet involves studying gene-environment interactions that mediate, change, or alter gene function (Liu et al., 2008). Environmental exposure to chemicals may influence biological system interaction. By studying chemical-biological system interaction insight may be provided into a chemical's mechanism of action, potential toxicity, and individual disease susceptibility.

ii. Mechanism of Action and Biomarkers

In pharmacology, the term Mechanism of Action (MOA) is referred to as the specific biochemical interaction in which a substance produces an effect or biomarker. Characterizing a chemical mechanism of action will allow for the identification of the chemical biomarkers (Heinzel et al., 2014). A biomarker is a measurable cellular, biochemical, or molecular alteration in human tissues, cells, or fluids that are indicative of biological processes, biological responses, or conditions (Mayeux, 2004). There are various types of biomarkers: genomic, transcriptomic proteomic, and metabolomic, which can be used as prognostic, predictive, or pharmacodynamic indicator. By studying biomarkers, and the

body's biological indicator of exposure, scientists are able to link chemical exposure to predictive health outcomes.

iii. Epigenetics

According to Liu et al. (2008), gene-environment interaction gives rise to epigenetic changes of the genome in response to a change in the environment. The gene-environment interaction may result in the alteration of gene operation, the formation of disease, and/or gene mutation.

iv. Chemical Structure

Chemical bonding results in a substance having unique physical and chemical properties. The number and type of chemical bonds determines the compound's structure, connectivity, and geometry. These unique properties have the ability to influence biological activity, specifically the toxicity of a compound (Vouk et al., 1987). Structure-activity relationships (SARs) have been used to describe a chemical's ability to induce adverse health effects (Vouk et al., 1987).

b. Burn Pits

Proper disposal of waste during in Afghanistan and Iraq have been essential in preventing unsanitary conditions and health hazards. When sanitary and waste management facilities are unavailable, military forces have relied on waste burning in the form of open-air burn pits as a method to reduce the volume of waste. Air emissions from open-air burn pits release

pollutants such as dioxins, particle matter (PM), polycyclic aromatic hydrocarbons (PAHs), and volatile organic compounds (VOCs) directly into the atmosphere (EPA, 2002). Current research conducted by the United States (US) Environmental Protection Agency (EPA) has found an increased risk of heart disease, asthma, and emphysema associated with the inhalation of air emissions from burning trash. In contrast, the health risks associated with military-related burn pit exposures have not been thoroughly assessed.

In response to service member concerns, an environmental health risk assessment was accomplished in 2007 in an attempt to characterize burn pit emissions. Initial screening samples, risk assessment models, and airborne concentration calculations performed were reported to the Department of Defense (DOD) Defense Health Board (DHB). The composite risk estimate performed compared burn pit air sampling result values to the one-year military exposure guidelines (MEGs), evaluated the severity of health risks, along with the probability and frequency of the exposure occurring. The risk estimate determined that exposures to volatile organic compounds (VOCs), polycyclic aromatics hydrocarbons (PAHs), dioxins, and furans were considered low (Weese, 2010). The health risk assessment rating of "low" was provided to commanders and the DHB in 2007.

The DHB determined that several problems existed with risk assessment, to include the infrequency of detection and sampling of burn pit emissions. Furthermore, the data presented to the DHB from Joint Base Balad (JBB) Iraq burn pit sampling employed quantitative screening using the Human Health

Risk Assessment Guidance for Superfund methodology outlined by the EPA. The EPA Methodology compared values designed for long-term exposure scenarios regarding a general population that included children and the elderly. Service member cancer risks were also estimated using theoretical or probabilistic cancer risk estimates and the service member's time on the base. The time periods used in the risk estimate included 24-hour days, seven days per week for twelve-month, four-month, and one-month exposure increments. The EPA method determined that both non-cancer and cancer risks were acceptable and considered safe per EPA classification (Weese, 2010).

According to the DHB, the report offered limited data examination and information on the potential effects of burn pit combustion exposures. Neither the amount, nor the type of material disposed of in the burn pits, was well controlled, defined, or characterized—thus resulting in the DHB believing that the burn pit emissions were not fully characterized. Fearing that burn pit combustion products pose an inhalation hazard that can potentially increase the long-term health risks for exposed service members, a need arose to characterize the gaps in procedure and determine risk-assessment practices, which could be effectively employed in austere and hostile environments (DHB, 2008).

As multiple complaints arose from veteran service members regarding burn pit exposures, so has the need to characterize the emissions and exposures from open air burn pits. In 2010, 300 veteran service members joined a class-action lawsuit against Kellogg Brown and Root (KBR), a

military contractor that operated several burn pits at bases in Iraq (NY Times, 2010). In response to the growing concern for service-members' exposure to harmful contaminants from open burn pits, and increases in respiratory illnesses in returning veterans, Congress and the Department of Veterans Affairs (VA) were directed to conduct studies which would determine the long-term health effects from open-air burn pit operations (IOM, 2011). At the same time, the Armed Forces Health Surveillance Center (AFHSC) and the Naval Health Research Center (NHRC) were tasked to conduct epidemiologic studies to determine any associations between burn pit emissions and exposure-related illnesses (Armed Forces Health Surveillance Center, Naval Health Research Center and US Army Public Health Command, 2010). Later, the VA commissioned the Institute of Medicine (IOM) to assess the long-term health effects from exposure to burn pits in Afghanistan and Iraq (IOM, 2011).

The IOM committee used the burn pit at Joint Base Balad (JBB) Iraq as the basis of their study to assess the long-term health effects from exposure to burn pits. To aid in the IOM study, the DOD provided raw air-sampling data, which were used to determine which chemicals were present at JBB, and which chemicals were present in ambient air. Based on this data, the committee found that levels of most pollutants at JBB were not higher than the levels measured at other polluted sites worldwide. However, IOM found insufficient evidence that prevented the committee from developing firm

conclusions regarding the long-term health effects caused by burn pit exposures.

c. Artificial Neural Networks and predictive modeling

An Artificial Neural Network (ANN) is a computational model based on the neural structure of a human brain. ANNs are comprised of interconnected processing elements or neurons that are trained to solve specific problems. As a nonlinear statistical data-modeling tool, ANNs establish relationships between an input and an output and models the result. Results from a study performed by Brouch (2014) provided the foundation that ANNs may have the potential to predict chemical-disease associations.

2. Problem Statement

The Comparative Toxicogenomics Database (CTD) is a publicly available database that integrates scientific literature to understand how environmental chemical exposures affect human health. The CTD organizes, integrates, and maintains data illustrating the interactions between chemicals and genes, chemicals and proteins, chemicals and disease, and genes and diseases (Davis et al., 2014). The organization, integration, and maintenance of chemical data by the CTD results in the data being classified as curated. However, current sampling of mock burn pit emissions identified several constituents that currently lack research and data. The lack of current research and data constitutes that these constituents be classified as uncurated. It is hypothesized that an ANN, coupled with current and relevant literature on uncurated burn pit constituents could map uncurated constituents to diseases across a wide variety of test species.

3. Research Objectives

a. Determine the effect of Hidden Layers (HL) on ANN performance for constituents relevant to burn pit emissions.

b. Determine the effect of Training, Validation, and Testing (TVT) ratios on ANN performance for constituents relevant to burn pit emissions.

c. Determine the effect of chemical structure on ANN performance for constituents relevant to burn pit emissions.

d. Identify key data gaps needed to advance ANN-based screening of constituents relevant to burn pit emissions.

4. Research Focus

The focus of this research is to predict the long-term health effects of uncurated burn pit emission constituents and verify ANN chemical-disease predictions with current and relevant literature. Using current literature, data and informational gaps on the uncurated burn pit constituents will be determined. The identified informational gaps will institute the need for further research in order to establish relationships between constituents, mechanisms of action, biomarkers, and disease.

5. Methodology

Utilizing the mock burn pit constituent analysis performed by Woodall (2012), the 47 constituents identified in the mock burn pit analysis were assessed using the CTD to verify curation status. Of the 47 constituents, ten constituents were determined to be uncurated. A curation of the ten uncurated chemicals was conducted to garner possible chemical-disease associations. Using the ANN established by Brouch (2014), the ANN was modified and tested using the ten uncurated constituents and the

disease associations curated during the literature review. Using the uncurated constituents, three tests were accomplished to determine the optimal ANN performance standards. The first test studied the effect of Hidden Layers (HL) on ANN performance. The second test was performed with three different TVT ratios, which were then compared to determine their effect on the ANN performance. The third test performed determined if chemicals with similar structures had an effect on the performance of the ANN. The culmination of all three tests established standard protocol for analyzing uncurated constituents.

6. Limitations

a. As the number of HL increased, the ANN response time slowed. This delay in response time indicated that the ANN had an optimal HL size. Five HL sizes were selected and tested. Three of the five tests demonstrated positive response times. Due to time limitations, only one of the three positive response times was selected and further evaluated.

b. In order to test and train the ANN, research was conducted to determine the current diseases associated with each constituent. If a constituent was found to have only two associated diseases then the model would be trained and tested to associate only two random diseases. Since uncurated constituents have minimal research accomplished for them, it is presumed that uncurated constituents may have more disease associations then those tested in this thesis.

c. This work did not track specific animal species disease-chemical associations. All known animal species disease-chemical associations were used as input

into the ANN regardless of their animal species pairing. However, animal species associations may be mapped to disease-chemical pairings using ANNs.

d. Relevant research on the uncurated constituents is limited to the availability of research, testing, and evaluation for each one of the uncurated constituents.

e. The CTD maintains limited data on chemical mixtures and the synergistic effects from chemical exposure.

7. Implications

Chemical exposure effects are determined through extensive and costly research, testing, and evaluation. The focus of this thesis is to refine a previous-established ANN to garner an alternative method for determining chemical exposure effects. This alternative may reduce the time and costs needed to research, test, and evaluate chemical hazards. The refined ANN may also serve as a decision support tool for military commanders, when faced with operations that involve the use of chemical hazards.

II. Literature Review

1. Relevant Research

a. The Relationship between chemicals and disease

i. Etiology:

This thesis focuses on establishing the need to understand how environmental agents influence disease. Through genomic and toxicological studies, scientists and researchers have established that chemicals have the ability to mutate, alter, and interact with genes often influencing gene expression and protein function (Lane, 2002). The effects of chemical compounds must be further characterized to understand the biochemical and genetic complexity that the chemical imposes on cells, tissues, organs, and overall human health.

In order to characterize a chemical's impact, the Environmental Protection Agency (EPA) uses chemical risk assessments to elucidate the human health and ecological risks from chemical hazards. As of 2015, the EPA reported approximately 85,000 chemicals listed on the Toxic Substance Control Act (TSCA) Inventory (EPA, 2015). Annually, the EPA receives 400 new Notice of Commencements (NOC) for chemicals that may modify the TSCA inventory. Due to the amount of NOCs received by the EPA, the EPA has identified data gaps in the evaluation of chemicals. These data gaps include; chemical use, exposure pathways, and toxicity data. The significance of the missing data prevents chemical risk assessments from being accomplished.

The data gaps identified by the EPA, and the absence of chemical risk assessments, established a need for alternate methods when performing chemical risk assessments. As an alternative method to understand how environmental chemicals affect human health, North Carolina State University (NCSU) developed the Comparative Toxicogenomics Database (CTD). The CTD commenced a curation or data collection of environmental chemicals (Mattingly et al., 2003). The CTD established a collaborative database with curated chemical data describing the relationships between chemicals, genes, proteins, and human diseases to advance the understanding of chemical effects on human health. The CTD integrates all data to facilitate the construction of chemical-gene-disease network and provide the groundwork for investigating the molecular basis of chemical-disease associations and toxicity (Davis et al., 2008).

Despite the extensive construction and curation of data provided by the CTD, many chemicals remain uncurated. Uncurated chemical data is absent of gene-interactions despite disease association, mechanisms, or biomarker associations. Due to the data gaps, relevant information is missing that would otherwise help understand the basis of disease. Despite containing uncurated data, the CTD contains the tools needed to generate testable hypotheses regarding the underlying etiology of chemical-disease relationships (Davis et al., 2015).

ii. Mechanisms of Action/Mode of Action and Epigenetics:

A Mode of Action (MoA) describes a cellular functional or anatomical change resulting from environmental exposure (FDA, 2011). In comparison, a Mechanism Of Action (MOA) is used in pharmacology to describe biochemical interaction at the molecular level. The Federal Food, Drug, and Cosmetic (FD&C) Act (2016) interprets chemical action through either chemical reaction or intermolecular forces or both. Chemical action either results in a bodily response at the cellular or molecular level or when the chemical binds with or modifies a molecular target or receptor.

Identifying the effect of exposure on human health is a major objective of biomedical research. In environmental health studies, it is recognized that environmental exposure could produce Deoxyribonucleic Acid (DNA) mutations. As a consequence of DNA mutation, chemical substances have been categorized according to their ability to alter DNA (Pulliero et al., 2015). A chemical's ability to alter DNA established a fundamental effort to determine risk assessment procedures, prevention, exposure reduction, and regulatory efforts.

Unlike the chemical substances that cause, or are suspected to cause, DNA mutation, there are chemical substances that cause gene expression and heritable change without changing or mutating the DNA sequence. The study of heritable change and gene expression is called epigenetics. Epigeneticist's investigate heritable changes in gene expression occurring through the mechanisms of DNA methylation, histone modification, and microRN

expression (Hou et al., 2012). In vitro, animal, and human studies performed by Baccarelli and Bollati (2009) have identified several classes of environmental chemicals that modify epigenetic mechanisms resulting in the identification of epigenetic mechanisms that may mediate specific mechanisms of toxicity and specific chemical response. Various chemical mechanism/mode of action are understood, others remain unidentified.

Baccarelli and Bollati (2009) reported that certain chemical exposures had altered epigenetic mechanisms, and that the same or similar epigenetic alterations were found in patients with the disease of concern or in diseased tissue. However, it was not determined whether the exposed individual's developed epigenetic alterations over time or which alterations increase the risk of developing disease. Forgoing the identification of these factors increased the difficulty in establishing the chemical-disease relationship between the chemical, the epigenetic change, and the presence of disease. At this time, very little is known about which epigenetic alterations are part of normal variability and which alterations are considered adverse. In order to understand epigenetic variability and the potential for chemicals to induce epigenetic modifications or alterations, an understanding of the chemical's mode of action (MoA)/mechanism of action (MOA) is needed.

A workshop held by the National Academy of Sciences' (NAS) Standing Committee on the Use of Emerging Science for Environmental Health Decisions discussed the role of chemically induced epigenetic changes in regulatory and policymaking decisions. The committee concluded that

epigenetic testing is not sufficiently validated for inclusion into any regulatory process. At this time, there is not a single/uniform test for epigenetic effects. Without understanding epigenetic effects, the patterns, variability, and long-term health effects relating to epigenetic changes are not well understood. As part of the committee's recommendations, there remains a need to establish a tiered epigenetic screening process to prioritize chemicals for further epigenetic analysis (Adler, 2010).

iii. Biomarkers:

Silins and Hogberg (2011) divide the use of biomonitoring into three classes that measure biological markers of exposure, effect, and susceptibility. Where, biomarkers of exposure measure the parent compound, the parent compound's metabolites, the biologically effective dose, and the biological effect (Silins and Hogberg, 2011; WHO, 2011). Silins and Hogberg (2011) identified biomarkers of effect as the cellular changes that alter expression of metabolic enzymes and disease development, at the same time the World Health Organization (WHO) (2011) identifies biomarkers of effect as structure alteration, altered function, or clinic disease. Lastly, the biomarker of susceptibility indicates the ability of an individual to respond to an environmental exposure (WHO, 2011).

Traditionally, the EPA and other federal agencies characterize environmental risk by extrapolating a chemical dose from in vitro and in vivo toxicological studies. According to Shatkin and Ranalli (2007), the use of biomonitoring will refine the risk characterization procedure by providing an

understanding of the fate and behavior of a chemical once inside the body. The use of biomonitoring may validate, improve, and alter future pharmacokinetic modeling, clinical applications, and toxicogenomic applications (Shatkin and Ranalli, 2007).

Currently, biomonitoring is used in both clinical and toxicogenomic applications. In clinical applications, biomonitoring is used to characterize exposures that occur in the occupational environment. It is believed that biomonitoring data can measure and assess environmental exposure trends, which assists with defining the relationship between exposure and disease. Biomonitoring may also provide an employee baseline prior to exposure.

In toxicogenomic applications, data from toxicogenomic studies are compared with in vitro cellular studies to identify chemical toxicity mechanisms. Utilizing exposure measurements, gene expression changes, and traditional toxicological markers, the identification of chemical toxicity mechanisms has been processed. The chemical toxicity mechanisms can then be used to generate biomarkers related to chemical exposure, chemical effect, or individual susceptibility (McHale et al., 2010). Similarly, the European Commission (EU) funded a research project on gene expression analysis and its use in relating biomonitoring to environmental carcinogenic exposures (Van Leeuwen et al., 2008). Van Leeuwen et al. (2008) consider the use of biological monitoring and biomonitoring as a way to profile gene expression in relation to human environmental exposures.

The EU research project found that gene expression profiles differed according to the population's environmental exposure. This finding resulted in the correlation between gene expression and the blood/urinary measures of biomarkers used to observe environmental carcinogen exposure. The EU project also provided evidence that exposure to environmental carcinogens affected the metabolism, stress response, signaling pathway, and the tumorigenesis of the studied genes. Overall, the EU is pressing for an increase in biomonitoring activities during the analysis of environmental health risks (Van Leeuwen et al., 2008).

According to Adelman (2005) the United States employs inadequate chemical testing methods to determine chemical risk. Both chemical risk assessments and toxicological studies face limitations due to the lack of chemical information. As a method to combat these limitations, the integration of genomics and toxicology is developing into a new research field call toxicogenomics. Toxicogenomics will help to identify biomarkers of exposure as well as relate disease to an environmental exposure. To promote the technological development to monitor gene, protein, and metabolite expression, the National Institute of Environment Health Science (NIEHS) has developed the National Center for Toxicogenomics. As of 2002, the Center was awaiting a tool to monitor the expression of thousands of genes, proteins, metabolites, and gene-environment interactions (Tennant, 2002).

iv. Chemical Properties

A method of assessing chemicals is through structure-activity relationships (SARs). SARs are designed to find the relationship between chemical structure, chemical structure related properties, and biological activity (OECD, 2016). SARs link chemical structure to a chemical property or a biological activity such as toxicity (OECD, 2016). According to the Organization for Economic Co-operation and Development (OECD) (2016), the theory behind SARs is that the activity of the chemical is found within the chemical's structure. Therefore, the structure of a chemical contains the features responsible for the chemical's physical, chemical, and biological properties (OECD, 2016). The biological activity of a compound may alter the chemical or physiological function of a cell, tissue, organ, or organism through the compound's physical and chemical makeup, concentration, and duration of exposure. OECD (2016) presumes that biological activities of a compound are governed by the compound's properties, which are determined by the compound's structure.

Abraham et al. (1989) considered the molecular weight and number of hydrogen bond donors and acceptors as physicochemical properties. Physicochemical properties predict a chemical's physical hazard, reactivity, and pharmacokinetics to include the chemical's absorption through exposure routes, the chemical's distribution in the body, and the chemical's metabolites. By understanding a chemical's pharmacokinetics and physicochemical properties, Abraham et al. (1989) identified the key properties responsible for

chemical absorbtion through various exposure routes. Molecular weight, the number of hydrogen-bond donors and acceptors, and $\log K_{ow}$ were identified as predictors of oral absorption. Chemical absorption through the skin depended upon the degree of hydrogen bonding. Unlike ingestion and cutaneous absorption, chemical inhalation prediction requires the use of vapor pressure, water solubility, and the chemical's reactivity.

Understanding the chemical characteristics that affect toxicity, Abraham et al. (1989) determined that a chemical with a large number of hydrogen bond donors had a reduced permeability rate. Chemicals with five or fewer hydrogen bond donors had an increase in permeability. When studying the molecular weight, Abraham et al. (1989) determined that lower molecular weight compounds maintained a lower lipophilicity, while higher molecular compounds resulted in higher lipophilicity. As a result of this study, Abraham et al. (1989) determined that higher molecular weight chemicals resulted in reduced blood brain barrier permeability (Abraham et al., 1989).

Furthering the Abraham et al. (1989) study, Lipinski et al. (1997) analyzed the physiochemical properties of 2,000-plus drugs to determine drug permeability and absorption potential. During their analysis, Lipinski et al. (1997) determined that permeability rates and absorption potential increased due to a compound's molecular weight, lipophilicity (expressed as logP), number of hydrogen bond donors, and the number of hydrogen bond acceptors. The Lipinski et al. (1997) findings are known as the Lipinski Rule of Five; which determines a compound's membrane permeability and ease of

absorption when the compound has a molecular weight less than 500, a lipophilicity less than five, the number of hydrogen-bond donors is fewer than five, and the number of hydrogen-bond acceptors is fewer than 10.

Using known toxic compounds, Struck et al. (2008) determined the biological activity toxicity profile for 50,000 toxic compounds for use as a toxicological classification guide. Struck et al. (2008), identified toxic compounds based on their structural properties. The structural properties identified as "toxicity properties" contain toxicity defining attributes identified as the compound's molecular weight, the number of hydrogen bond donors and acceptors, and functional groups (Struck et al., 2008). Struck et al. 2008 identified that the use of SARs for predicting toxicity is limited due to the complexity of the structure of toxic biological macromolecules, the variability of metabolic pathways, toxicity differences among animal and plant species, the effects of substituents on the reactivity of the core chemical, and the effect of steric properties. However, some researchers have developed SARs based on key structure components.

b. Burn pits

i. Burn Pit Emissions:

Due to the limited nature of open air burn pit emission studies, and the difficulty of assessing, characterizing, and quantifying burn pit emissions and associated exposures, Lemiux et al. (2004) identified that emissions from burning waste varied from source to source. Lemiux et al. (2004) identified

that the fuel composition, fuel heating value, bulk density, oxygen transport, and combustion varied from source to source.

Woodall et al. (2012) attempted to characterize burn pit emissions during small-scale emissions testing of simulated military deployment waste. Woodall et al. (2012) identified burn pit constituents using a representative collection of military waste based on expert knowledge of DOD personnel. The waste stream sampled consisted of large amounts of styrofoam, electronics, packaging materials, construction materials, food waste, canvas material, Meals Ready to Eat (MRE) waste, and plastic water bottles. The analysis of the emissions testing concluded that 47 volatile organic compounds (VOCs) were present in the burn pit emissions. Ten of the 47 VOCs lacked constructed curated chemical-disease relationships by the CTD, thus becoming the uncurated chemical constituents tested within this thesis.

Aurell et al. (2012), further characterized burn pit emissions to assess potential inhalational exposures. Aurell et al. (2012) collected 41 Semi-VOCs (SVOCs), $PM_{2.5}$ by filter, and VOC samples. 21 VOCs were found to be present in the Aurell et al. (2012) burn pit emissions. Three of the 21 VOCs lacked constructed chemical-disease relationships by the CTD. The Woodall et al. (2012) and the Aurell et al. (2012) study found two of the same uncurated constituents: propene and vinyl acetate. 15 of the 21 VOCs identified in the Aurell et al. (2012) study are listed as EPA hazardous air pollutants. The Aurell et al. (2012) study also identified the following

emissions: phenanthrene, naphthalene, lead, iron, copper, chromium, arsenic, nickel, and cadmium.

ii. The Relationship between Uncurated Burn Pit Constituents and Disease

The following research was accomplished to identify the relationship between the ten uncurated burn pit constituents identified by Woodall et al. (2012) and their known disease associations. In order to understand the relationship between uncurated constituent and disease, the mechanism or mode of action was also identified for each uncurated chemical. The qualitative impacts of the associated diseases, mechanisms and modes of action are not fully defined and exclude the quantitative differences between test species. Additionally, the mechanism or modes of action that cause disease are not considered a constant property of the chemical. The mechanism or mode of action that causes disease are subject to variation between species and may change with chemical concentration and duration of exposure (Nendza and Wenzel, 2006).

The physio-chemical properties of chemical compounds are toxic in different ways due to the compound's interactions at the biomolecular level. Nendza and Wenzel (2006) identified that physio-chemical properties also determine the transport and interaction of the compound with biomolecular targets. Partitioning of the chemical across a membrane often resulted in non-specific toxicity, while chemical interaction between compounds with specific targets produced an increase in toxicity (Nendza and Wenzel, 2006). Nendza and Wenzel (2006) also determined that chemical compounds did not

maintain the same mechanism of action during cross species analysis. The change in the mechanism of action across species was determined to occur from the change in the abundance of mechanism specific targets and the endpoint measured by each experiment.

Chemical compounds are also hypothesized to interact with several targets to varying extents in different species. These interactions between the chemical and the varying targets may result in multiple concurrent effects. 20% of the chemicals tested by Nendza and Wenzel (2006) demonstrated two concurrent mechanisms of disease, and 5% of the chemicals tested revealed three concurrent mechanisms of disease.

The following diseases, mechanisms or modes of action for the ten uncurated burn bit constituents are based on the best available research. The MoA/MOA for each constituent was selected using metabolism and pharmacokinetic studies.

1. 4-Ethyltoluene

   4-Ethyltoluene is an active sister chromatid exchange agent in vitro balb mice bone marrow cells (Janik-Spiechowicz and Wyszynska (1998). Sister chromatid exchanges (SCEs) involve the breakage of both DNA strands, followed by an exchange of the whole DNA strand. SCEs have been correlated with recombinational repair and the induction of point mutations, gene amplification and cytotoxicity (CRIOS, 2008). According to the National Library of Medicine (2013), point mutations are associated

with the development of cystic fibrosis. The Cancer.Net Editorial Board (2012) found that point mutations have been associated with cancer.

Swiercz et al. (2000) found that in vivo rat repeated-dose inhalational toxicity studies to 4-ethyltoluene resulted in elevated serum levels of gamma-glutamyl transferase (GGT); a biomarker of oxidative stress. Oxidative stress may damage cells by damaging proteins, lipids, and DNA which can disrupt the cellular signaling mechanisms. Slupphaug et al. (2007) identified DNA Damage caused by oxidative stress may prevent proper base pairing resulting in strand breaks. In humans, oxidative stress is hypothesized to be involved in the development of neurodegenerative diseases (Patel and Chu, 2011), cancer (Halliwell, 2007), atherosclerosis (Bonomini et al., 2008), heart failure (Singh et al., 1995) myocardial infarction (Ramond et al., 2011), and chronic fatigue syndrome (Kennedy et al., 2005).

Swiercz et al. (2000) also identified the concentration-related increase in lactate dehydrogenase (LDH) with in vivo rat repeated-dose inhalational toxicty studies to 4-ethyltoluene. LDH is medically significant due to its release when tissue damage has occurred. As a biomarker of oxidative stress, LDH is a commonly linked to heart failure, liver disease, cancer (Gersten, 2014), lung disease (Drent et al., 1996), and hypothyroidism (McGrowder et al., 2011). Yorifuji et al. (2012) identified a prevalence of respiratory and mucocutaneous zone irritation when

exposed to VOCs containing 4-ethyltoluene. 4-Ethyltoluene mechanisms stated above are consistent with the irritant properties of VOCs.

2. Benzanthrone

Singh et al. (2003) discovered that benzanthrone could induce lipid peroxidation in vivo guinea pig comparative safety assessments. Lipid peroxidation is one of the main molecular mechanisms involved in the oxidative damage to cellular structures and the toxicity process that leads to cellular death (Repetto et al., 2012). According to TOXNET (2003), benzanthrone induced toxicity is suggested to take place in areas of the cell where there is an increase in the production of lipid peroxides. These areas include the nuclear, mitochondrial, and microsomal membranes.

Further evaluations revealed enhanced oxidative stress, upregulation of prominent inflammatory markers, and DNA damage coinciding with benzanthrone exposures. Tewari et al. (2015) found that the administration of benzanthrone in vivo studies with mice also induced enzyme activation that consisted of mechanistic pathways involved in inflammatory manifestations, which suggests that benzanthrone is an immunotoxin agent. Upon exposure to light benzanthrone generates active oxygen species. The National Research Council (1999) hypothesized that the active oxygen generation of benzanthrone might be responsible for the photo-contact dermatitis experienced in humans. The oxidative stress mechanism associated with benzanthrone is consistent

with the skin disorders documented with contact exposures to benzanthrone.

Sidhu et al. (2005) identified benzanthrone as a suspected human Endocrine disrupting chemical (EDC) based on the structual similarities to both known and suspected EDCs. De Coster and Van Larebeke (2012) discerned that EDCs mimic or block the transcriptional activation elicited by naturally circulating steroid hormones by binding to steroid hormone receptors. EDCs increase or block the metabolism of naturally occurring hormones by activating or antagonizing estrogen, androgen, and thyroid hormone receptors. Tabb and Blumber (2013) observed EDCs effect on the genome, DNA, and lipid metabolism.

3. Benzyl Chloride

Evaluation of benzyl chloride by TOXNET (2014) identified benzyl chloride as causing an increase in the alkylation of DNA resulting in alkylation lesions in DNA and RNA in vivo mice studies. Drabløs et al. (2004) demonstrated how alkylation lesions may result in genotoxicity and cytotoxicity, eventually resulting in cytotoxic double-strand breaks of the DNA. Alkylating agents, like benzyl chloride, inhibit the transcription of DNA into RNA, thus stopping protein synthesis, which triggers apoptosis. Alkylating agents also substitute alkyl groups for hydrogen atoms in the DNA sequence, resulting in the formation of cross-links in the DNA chain. The National Library of Medicine (NLM) (2015) identified that cross links in the DNA sequence can result in cytotoxic,

mutagenic, and carcinogenic effects. Typically, alkylating agents primarily affect hematopoietic cells, reproductive cells, endothelial cells, bone marrow, and the cells of the gastrointestinal tract. The NLM (2015) also associated the common side effects of alkylating agents. These side effects include anemia, pancytopenia, amenorrhea, impaired spermatogenesis, intestinal mucosal damage, alopecia, and an increased risk of malignancy. Benzyl Chloride animal carcinogenicity is consistent with the alkylation of DNA mechanisms.

Similar to benzanthrone, benzyl chloride also induces lipid peroxidation during in vitro rat liver testing (EC, 2014). Benzyl chloride also produced SCE with point mutations in vitro hamster studies (EC, 2014). Disparate from the other uncurated burn pit chemicals, benzyl chloride is associated with a dose-dependent increase in the number of chromatid aberrations in rat epithelial cells (EU, 2014). The chromatid aberrations created by benzyl chloride produce breaks and gaps in the chromatids during and after replication. Chromatid aberrations often result in chromosomal or genome mutations.

Benzanthrone is a Polycyclic Aromatic Hydrocarbon (PAH). Abdel-Shafy and Mansour (2015) describe PAHs as highly lipid soluble in mammalian gastrointestional tracts. Abdel-Shafy and Mansour (2015) further describe both acute and chronic effects of PAH exposure with the caveat that short-term human health effect information of PAHs is not clear. High levels of PAH exposures have resulted in eye irritation,

nausea, vomiting, diarrhea, and confusion in humans (Abdel-Shafy and Mansour, 2015). Long-term human exposures to PAHs have resulted in decreased immune function, cataracts, kidney damage, liver damage, breathing problems, asthma-like symptoms, lung function abnormalities, skin redness, and skin inflammation (Abdel-Shafy and Mansour, 2015).

Irigaray and Belpomme (2009) suggest that PAHs with low molecular weights and three to four rings are non-genotoxic carcinogen promoters causing the expansion of initiated cells. Non-genotoxic promoters are also suggested to cause preneoplastic cells to escape cellular growth control mechanisms. Irigaray and Belpomme (2009) also suggest that PAHs are cocarcinogens, whereas, cocarcinogens activate carcinogens, and may enhance the carcinogens effects. The mechanisms by which cocarcinogens activate, enhance, or effect carcinogens are through the depletion of detoxifying proteins, inhibit enzymatic activity, inhibit DNA repair enzymes, and/or activate procarcinogenic material into carcinogenic material.

4. n-Heptane

The irritation and central nervous effects occurring from n-heptane exposures are linked to the lipophilic properties of n-heptane (MAK Value Documentation, 1998). Szutowski (2009) suggests that cytochrome P450 produces n-heptane metabolites which modulate the biotransformation of n-heptane, resulting in the oxidation of n-heptane during in vitro rat studies. Oxidation in the body can damage cell membranes, cellular

proteins, lipids, and DNA. Oxidative DNA damage may result in the production of single or double stranded DNA breaks, base modifications, or rearrangements. Due to the limited research available, n-heptane's mechanisms of disease could not be compared to n-heptane symptomology.

5. n-Octane

Similar to n-heptane, n-octane is also metabolized by cytochrome P450 and undergoes oxidation. An in vivo rat study by Pandya et al. (1982) cited in TOXNET (2014) observed the increase in liver and spleen alkaline phosphatase (ALP) when exposed to n-octane. Increased ALP activity is associated with hepatobiliary and bone diseases. Tietz (1999) associated elevated ALP levels with disorders of the skeletal system, hyperparathyroidism, osteomalacia, fractures, and malignant tumors. Khan et al. (1980) linked n-octane to lipid peroxidation induction. Yorifuji et al. (2012) identified a prevalence of respiratory and mucocutaneous zone irritation when exposed to VOCs containing n-octane. n-Octane mechanisms stated above are not consistent with the irritant properties of VOCs.

6. Propene

Propene is expected to undergo hydration becoming an alcohol, and is later excreted as a conjugated alcohol or propionic acid. Similar hydroxylate chemical reactions involve cytochrome P450. TOXNET (2005) hypothesizes that propene may be metabolized by cytochrome

P450. Due to the limited research available, propene's mechanisms of disease could not be compared to propene symptomology.

7. Salicylaldehyde

Freda (1970) demonstrated salicylaldehyde's enzyme inhibition effects in vitro bovine testing. Enzyme inhibitors alter the catalytic action of enzymes causing a delay or discontinuance of catalysis. Similar to benzanthrone, the EPA classified salicylaldehyde as an EDC. De Coster and Van Larebeke (2012) suggest that EDCs contribute to cancer, diabetes, obesity, metabolic syndrome, and infertility. EDCs may act through classical nuclear receptors, but also through estrogen-related receptors, resulting in enzyme activation or modulation. De Coster and Van Larebeke (2012) proposed that EDCs interfere with feedback regulation, neuroendocrine cells, DNA methylation, which results in histone modifications, while Tabb and Blumber (2013) suggests that EDCs activate mitogen-activated protein kinase. Due to the limited research available, salicylaldehyde's mechanisms of disease could not be compared to salicylaldehyde symptomology.

8. Tetrahydrofuran

The mode of action and the biotransformation mechanisms of Tetrahydrofuran (THF) are not well understood. The Proposal for Harmonised Classification and Labeling (2009) hypothesized that tetrahydrofuran undergoes an alpha-hydroxylation, followed by a subsequent ring opening. This process is suspected to produce a

hepatotoxic aldehyde. The Proposal for Harmonised Classification and Labeling (2009) also hypothesized that the oxidation of the hydroxyl group before the ring opening occurs, leading to the formation of a gamma-butyrolactone; a potential neurotoxic and a gamma-hydroxibutyric acid.

The European Chemical Agency (ECHA) (2010) suggests that tetrahydrofuran is capable of inhibiting cytochrome P450 in vitro rat biotransformation studies. This inhibition of cytochrome P450 may cause the formation of peroxides and formaldehyde to occur. The Proposal for Harmonised Classification and Labeling and ECHA also suggest that tetrahydrofuran oxidative metabolism occurs due to cytochrome P450. However, the enzymes responsible for Tetrahydrofuran metabolism are not yet known.

Mode of Action in vivo rat research conducted by ECHA suggests that the tetrahydrofuran is responsible for liver toxicity while the metabolites of tetrahydrofuran are responsible for neurological effects. In vivo tetrahydrofuran studies also produced mitogenic effects in mice tissue, yet tetrahydrofuran demonstrated the ability to modify drug absorption and metabolism mechanisms, which may give rise to chemical distinctive toxicities.

Chhabra et al. (1998), and the U.S. Department of Health and Human Services (1998) as cited in TOXNET (2011), identified tetrahydrofuran as an animal carcinogen during in vivo rat and mouse inhalational studies.

Tetrahydrofuran demonstrated positive results for liver adenoma/carcinoma, and renal tubule adenoma/carcinoma, in both rat and mouse test subjects. These studies indicate that the oxidative stress and the metabolism of tetrahydrofuran are realistic mechanisms of disease for tetrahydrofuran exposures.

9. Triphenylene

Triphenylene is a Polycyclic Aromatic Hydrocarbon (PAH). Historically, the toxic effects of highly lipid soluble polycyclic aromatic hydrocarbons were not well documented (Sikkema et al., 1995). However, Sikkem et al. (1995) concluded that the specific toxicity of certain PAHs, when performed with a liposomal model, show that the absence of mass transfer limitation affects the PAHs energy transduction across biological membranes.

Irigaray and Belpomme (2009) suggest that PAHs with low molecular weights and three to four rings are non-genotoxic carcinogen promoters causing the expansion of initiated cells. Non-genotoxic promoters are also suggested to cause preneoplastic cells to escape cellular growth control mechanisms. Irigaray and Belpomme (2009) also suggest that PAHs are cocarcinogens, whereas, cocarcinogens activate carcinogens, and may enhance the carcinogens effects. The mechanisms by which cocarcinogens activate, enhance, or effect carcinogens are through the depletion of detoxifying proteins, inhibit enzymatic activity, inhibit DNA repair enzymes, and/or activate procarcinogenic material into carcinogenic

material. Due to the limited research available, triphenylene's mechanisms of disease could not be compared to triphenylene symptomology.

10. Vinyl Acetate

Similar to 4-ethyltoluene, vinyl acetate also induces SCE in mammalian in vitro studies. Unlike 4-ethyltoluene, vinyl acetate induced structural chromosome aberrations which further induce DNA cross-links mammalian in vitro studies. The Proposal for Harmonised Classification and Labeling (2010) identified that vinyl acetate induced changes causing fatty degeneration of hepatic parenchyma, and the proliferation and extension of smooth endoplasmic reticulum. Vinyl acetate also caused cell proliferation, cytotoxic effects, and mitotic inhibition during in vivo rat studies. Acetaldehyde, a metabolite of vinyl acetate through esterase-mediated metabolism, exhibited genotoxicity, induced DNA protein crosslinking and chromosomal damage in mammalian in vitro testing.

A risk assessment accomplished for the European Commission (EC) (2008) on vinyl acetate, identified vinyl acetate as genotoxic having a threshold mode of carcinogenic action in vivo rat studies. Genotoxic carcinogens damage DNA through covalently binding to the DNA. The DNA binding can occur as a direct binding to the DNA, after enzymatic activation, or by insertion into the DNA double helix (Van Delft et al., 2004). DNA damage can result in the dysfunction of the cell cycle, DNA repair, and lead to apoptosis.

The 2008 European Commission Risk Assessment identified vinyl acetate as inducing cellular proliferation at high levels during cancer studies. During the cancer studies, EC researchers assumed that cytotoxicity was the underlying mode of carcinogenesis. In vivo rat studies vinyl acetate exposures resulted in an increase in cell proliferation due to mitogenic actions, which lead to tumor formation. The European Commission (2008) presently accepts that acetaldehyde plays a critical role in the tumorigenicity of vinyl acetate, and suggests that the hydrolysis product of acetaldehyde is the active carcinogenic metabolite of vinyl acetate. Vinyl acetate symptomology is consistent with the established mechanisms of disease.

c. Using ANNs for predictive modeling

An artificial neural network (ANN) is a computational empirical model that is modeled after biological neurons. ANNs consist of inputs, weights, and mathematical functions in order to detect complex nonlinear relationships between dependent and independent variables in a given data set. Adjusting the weight of an ANN symbolizes how the model can be trained in order to obtain an output for a set of given inputs. Since the first use of ANNs in 1943 (Gershenson, 2016), ANNs have been used to map associations when the data contains variables that are vague, or difficult to describe. ANNs are currently used in diagnostic systems, biochemical analysis, image analysis, and drug development.

The pharmaceutical industry is using ANNs to associate subsets of physiochemical descriptors with biological activity profiles to provide an

understanding of how biological activity is elicited by a chemical structure. Llewellyn (2007) used this approach to predict drug likeness and toxicological effects. Unlike the pharmaceutical industry, the medical industry according to Rae et al. (1999) applied the use of ANNs as a decision support tool to help clinicians identify populations who are at an increased risk for developing specific diseases. The medical use of ANNs demands extensive development into the type and number of parameters used by the ANN—specifically the number of neurons in each layer and the applied learning algorithm.

Other medical uses of ANNs have occurred when Sheppard et al. (1999) used neural networks to identify potential patients at risk of developing cytomegalovirus disease. Santos-Garcia et al. (2004) also used ANNs to predict patients' morbidity from cardiorespiratory failure after non-small cell lung cancer pulmonary resection. Tseng et al. (2013) examined patient risk factors associated with hip fractures using ANNs. Alizadeh et al. (2015) created an ANN that accurately predicted the diagnosis of asthma.

d. Using ANNs to link chemicals and disease

ANNs have established their usefulness in medical and pharmaceutical prediction. However, research is limited to demonstrate how ANNs model chemical-disease association. Most often the research on the prediction of chemical-disease association involves the use of ANN Quantitative Structure Toxicity Relationship (QSTR) and an ANN Quantitative Structure-Activity Relationship (QSAR). ANN QSTRs are often used to map the relationship between a chemical's molecular descriptors and its toxicological activity, while

QSARs correlate physiochemical parameters to either a chemical or biological activity. Cheng and Sutariya (2012) describe the use of ANN QSTR in the study of pharmacotoxicology, and the use ANN QSAR as a screening approach during the drug discovery process.

Modeling drug response, Larder et al. (2007) demonstrated the use of an ANN developed to predict responses to antiretroviral therapy. Larder et al. (2007) identified that their developed ANN is limited by its capacity to only predict response to drugs that are included in the training dataset; thus, creating the need for further training of their model. Sibanda and Pretorius (2012) detailed the use of ANNs in the war against the Human Immunodeficiency Virus (HIV). As cited by Sibanda and Pretorius (2012), Dechao Wang et al. (2009**)** developed an artificial neural network (ANN) that predicted the virological response to HIV drug therapy. Similarly, Agatonovic-Kustrin and Beresford (1999) showed that their ANN model demonstrated better fitting and better predicting abilities in their investigation of the effects of pharmaceuticals. The ANN developed by Agatonovic-Kustrin and Beresford (1999) produced correlation values indicating strong association between the testing and training data. The high correlation values demonstrated that the ANN was able to predict pharmacokinetic and pharmacodynamic (PK-PD) relationships.

In the attempt to understand compound specific toxicity, Vracko et al. (1999) demonstrated the difficulty of using ANNs to predict toxic effects. Vracko et al. (1999) displayed the complications of predicting specific toxicities for 41 benzene analogues. However, the Vracko et al. (1999) ANN established that ANNs were

better capable of predicting toxicity classes versus predicting compound specific toxicity. Villemin et al. (1994) used an ANN to predict the carcinogenicity of 94 PAHs. The Villemin et al. (1994) ANN correctly associated and classified 81 of the 94 PAHs according to their activity. Gini et al. (1999) identified that molecules with identical or similar chemical descriptors differ in toxicity. The toxicity difference occurred due to the diversity in test animal metabolism. The ANN used by Gini et al. (1999) was not able to distinguish the toxicity between certain chemicals with similar descriptors, resulting in the need for further knowledge of the bioprocesses involved in chemical metabolism and the structural features of the chemical that characterize the chemical's specific mechanism of action.

2. Summary

The articles reviewed and discussed in the above literature review represent a small research sample related to the key concepts used to develop this thesis. The purpose of this literature review was intended to provide a general familiarization with current and relevant research regarding chemical-disease relationships, burn pits, and artificial neural networks. As a result of current and relevant research, the use of ANNs as predictors of disease have been used and verified in various capacities. However, the DOD as an aid in predictive applications has not employed the use of ANNs. Identifying burn pit constituents, their physicochemical properties, their anticipated acute health effects, and chronic health risks can be characterized using an ANN. Existing data gaps between chemical relationships, biomarkers, and disease will be identified with the expectation that further research will occur. While the goal

of the analysis was to provide a decision-support tool, the risk-assessment related burn pit exposures would aid in the anticipation, recognition, prevention, mitigation, control, communication, and documentation of future exposures.

III. Methodology

1. Data

a. Training Data

The data used to train the Brouch MATLAB® ANN was obtained and used previously by Brouch (2014). Training input data was originally obtained from curated Comparative Toxicogenomics Database **(**CTD) data. The chemicals used in the Brouch MATLAB® ANN were selected randomly from curated CTD data by Brouch (2014).

b. Testing Data

The data used to test the Brouch MATLAB® ANN was obtained from Woodall's (2012) burn pit constituent air sampling and thesis. Woodall (2012) identified 47 VOCs in the burn pit air sampling results. Each VOC was selectively compared to curated CTD data to verify the constituent's curation status. Ten of the 47 VOCs were not curated by the CTD. The ten VOCs were then curated by the author to identify the following: chemical structure, molecular weight, the number of hydrogen bond acceptors, the number of hydrogen bond donors, mechanism/mode of action, and related associated diseases.

c. Constituents

Table 1 displays the selected ten uncurated constituents. Table 1 lists the constituent name, synonyms, chemical abstracts service (CAS) number, molecular weight, structure diagram, associated diseases, and associated mechanism/mode of action. Chemical-associated diseases were selected based on toxicity studies

using only the specified chemical and no chemical mixtures from peer-reviewed literature.

i. 4-Ethyltoluene

4-Ethyltoluene is a benzene derivative. At this time, research is limited on 4-ethyltoluene and its effects (PubChem, 2015). A study performed by Yorifuji et al. (2012) identified 4-ethyltoluene as a VOC constituent being released from a plastic reprocessing factory in Japan. Based on the study performed by Yorifuji et al. (2012), the researchers found a prevalence of mucocutaneous and respiratory symptoms in residents closest to the plastics factory. Other symptoms experienced during this study include sore throat, eye itch, eye discharge, eczema, and sputum (Yorifuji et al., 2012).

ii. Benzanthrone

Benzanthrone is an industrial chemical that can be found in dye intermediates and as a product of fossil fuel combustion (TOXNET, 2003). Benzanthrone has been detected in particulates from diesel engines' exhaust, municipal waste incinerator ash, and wood and coal smoke. Due to the vapor pressure of benzanthrone, it will exist in both the vapor and particulate phases in the atmosphere. Exposure to benzanthrone can occur through dermal contact, inhalation, and ingestion. Skin disorders due to benzanthrone are more frequent in warm seasons and are significantly aggravated by heat and light (Encylopedia of Occupational Health and Safety, 1971)

Chronic exposures to benzanthrone have been documented to cause loss of appetite resulting in weight loss, intolerance of fatty foods, liver impairment, and gastritis. Workers reported general fatigue and weakness along with complaints of decreased sexual potency, increased neurasthenic syndrome, accelerated pulse, and reduced blood pressure. Contact exposures to benzanthrone have resulted in skin sensitization, eczema, erythema, dermatitis, and skin pigmentation. A study performed by Sidhu et al. (2005), identified benzanthrone as a suspected Endocrine Disrupting Chemical (EDC).

iii. Benzyl Chloride

Benzyl chloride is used as a chemical intermediate in the manufacture of dyes, pharmaceutical products, photographic developer, perfume, flavor products, synthetic tannins, pesticides, and petrol (TOXNET, 2015). Acute occupational exposures to benzyl chloride have resulted in respiratory tract, skin, eye, pulmonary, and mucous membrane irritation. Other symptomatology includes severe irritation of the upper respiratory tract with coughing, conjunctivitis, dizziness, weakness, headache, eyelid and finger tremors, increased bilirubin levels in blood, and a decrease in number of leukocytes. Lung damage, pulmonary edema, permanent eye damage, and CNS depression are suspected after severe benzyl chloride exposure (TOXNET, 2015).

Animal data indicate that long-term exposure to benzyl chloride showed an increase in the incidence of benign and malignant tumors, along with an

increase in thyroid tumors in female rats. At this time, there is limited human evidence for the carcinogenicity of benzyl chloride. Based on the coupling of animal studies and the insufficient and inadequate data on human studies, the EPA (2000) has classified benzyl chloride as a Group B2, probable human carcinogen. American Conference of Governmental Industrial Hygienists (ACGIH) classified benzyl chloride as a confirmed animal carcinogen with unknown relevance to humans (HSDB, 2015)

The EPA (2000) detected benzyl chloride in air emissions from the burning of polyvinyl chloride, neoprene, and urethane foam compounds. The EPA (2000) reported acute exposure to high concentrations of benzyl chloride causing central nervous system impairment including dizziness, headaches, weakness, and fatigue. Eye contact with benzyl chloride may result in permanent eye damage, while ingestion of benzyl chloride may cause mouth, throat, and gastro-intestinal tract burns that result in nausea, vomiting, cramps, and diarrhea (EPA, 2000).

The chronic effect of benzyl chloride exposure in humans has not been studied. However, chronic animal exposure to benzyl chloride suggests that benzyl chloride specifically targets the stomach and heart. At this time, there are no studies indicating human developmental or human reproductive effects. However, a rat in vivo study demonstrated an increase in embryonal mortality, along with developmental retardation in the offspring of exposed rats.

iv. n-Heptane

n-Heptane is used as an industrial solvent, anesthetic, laboratory reagent, in the petroleum refining process, and as a paint/coating additive (NJ DOH, 2007). Consumer use of n-heptane includes adhesives, sealants, automotive care products, ink, toner, and plastic/rubber products (PubChem, 2015). Acute exposure to n-heptane has been documented to cause eye, nose, and throat irritation, headache, lightheadedness, dizziness, lack of coordination, loss of consciousness, and loss of appetite. Chronic exposure to n-Heptane included skin rash, skin dryness, and skin redness. Limited research suggests that n-heptane may affect the central nervous system to include symptoms of memory loss, withdrawal, irritability, fatigue, sleep disturbances, and extremity weakness. Historically, the EPA (1989) did not classify n-heptane as a carcinogen and could not verify n-Heptane's ability to cause reproductive hazards.

v. n-Octane

n-Octane is used as a fuel additive, paint/coating additive and as a solvent (HSDB, 2014). A study performed by Yorifuji et al. (2012), identified n-octane as a VOC constituent being released from a plastic reprocessing factory in Japan. Based on the study performed by Yorifuji et al. (2012) a prevalence of mucocutaneous and respiratory symptoms in residents closest to the plastics factory was found and associated with n-octane exposures. Specifically, the symptoms of sore throat, eye itch, eye discharge, eczema, and sputum were increased among the residents (Yorifuji et al., 2012). Human

exposure studies to n-octane, have determined that n-octane exposure can cause giddiness, vertigo, headaches, stupor, epileptic seizures, respiratory tract irritation, visceral damage, chemical pneumonitis, pulmonary edema, skin blisters, and hemorrhaging (HSDB, 2014).

vi. Propene

Propene is used in plastics, carpet fibers, as a fuel additive, chemical intermediate, aerosol propellant, and in medication (HSDB, 2014). Acute exposures to propene have been documented to cause dizziness, lightheadedness, and loss of consciousness. Chronic exposures to propene include liver damage and irregular heart-beat. At this time, there is inadequate evidence and testing establishing propene's carcinogenicity. Propene has not been tested to verify reproductive hazards (NJ DOH, 2004).

vii. Salicylaldehyde

Salicylaldehyde is a chemical reagent and can be found in perfumes, fumigants, gasoline, flavor ingredients, and in medicinal chemicals (HSDB, 2003). A study performed by Sidhu et al. (2005) identified salicylaldehyde as a suspected Endocrine Disrupting Chemical (EDC). The EPA (2015) also considers salicylaldehyde an EDC.

viii. Tetrahydrofuran

Tetrahydrofuran is used as a reaction medium, reagent, and solvent in various operations that involve printing inks, adhesives, lacquers, coatings, fuels, pharmaceuticals, perfumes, insecticides, resins, vinyl, polymers, and cellophane. Data on Tetrahydrofuran toxicity is limited. TOXNET (2015)

identified several animal studies that correlated tetrahydrofuran exposure to skin and mucous membrane irritation, respiratory tract irritation, liver damage, kidney damage, lung damage, and gastro-intestinal tract inflammation. Similar to benzyl chloride, the ACGIH also confirmed tetrahydrofuran as an animal carcinogen with unknown relevance to humans (HSDB, 2015).

ix. Triphenylene

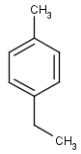
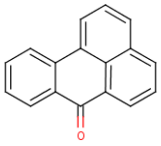
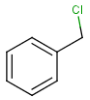
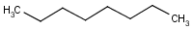
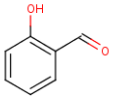
Triphenylene research and studies are rare in nature. Two mutagenicity studies were located and examined for relevance. Both studies identified triphenylene as causing mutagenicity with positive results in the AMES tests with and without S9 fraction from rat livers (CCRIS, 1985). The AMES test is used to determine if a chemical can induce mutations in DNA—thus establishing a chemical's mutagenicity (DeStasio, 2015).

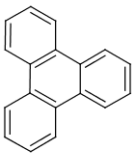
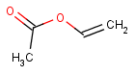
x. Vinyl Acetate

Vinyl acetate is used in plastics, films, lacquers, food packaging, food starches, polyvinyl emulsions, resins, coatings, paints, sealants, construction products, carpet backing, chewing gum, tablet coatings, acrylic fibers, air sprays, textiles, paper products, and laminates. A study by Budinsky et al. (2013) of vinyl acetate monomers (VAM) identified that VAM exposure produced nasal tumors in rats. According to studies compiled from the Hazardous Substances Data Bank (HSDB), there is a lack of evidence confirming vinyl acetate as a human carcinogen. However, vinyl acetate is shown to be genotoxic in human cells via in vitro studies and genotoxic to animals via in vivo studies; these findings resulted in the ACGIH

categorization of vinyl acetate as an animal carcinogen with unknown relevance to humans (HSDB, 2011).

Table 1: Chemical Information

Constituent Name	Molecular Formula	Chemical Structure	Known Disease Associations	Mechanism/Mode of Action
4-Ethyltoluene	C_9H_{12}		**Rat:** Ataxia (ChemIDplus) Convulsions (ChemIDplus) Bronchitis (TOXLINE) Pneumonia (TOXLINE) Perivascular Lymphoid Infiltrations (PLI) (TOXLINE)	Sister Chromatid Exchange (SCE) Gamma-Glutamyl Transferase (GGT) Lactic Acid Dehydrogenase (LDH)
Benzanthrone	$C_{17}H_{10}O$		**Guinea Pigs:** Skin Sensitization Reactions (HSDB) Liver Necrosis (HSDB) Kidney/Bladder Lesions (HSDB) (Sex Unspecified)	Lipid Peroxidation Oxidative Stress Immunotoxin Endocrine Disruption (EDC)
Benzyl Chloride	C_7H_7Cl		**Cat:** Pulmonary Edema (ChemIDplus) Corneal Damage (ChemIDplus) **Mice:** Respiratory Depression (ChemIDplus) **Rat:** Respiratory Depression (ChemIDplus) **Multi Species:** Neoplasms (HSDB)	Alkylating Agent Lipid Peroxidation Chromatid Aberration
n-Heptane	C_7H_{16}		**Human:** CNS Depression (HSDB) Chemical Pneumonia (HSDB) Dermatitis (HSDB) Cardiac Sensitizer (HSDB) Hallucinations (ChemIDplus)	Cytochrome P450 Oxidation
n-Octane	C_8H_{18}		**Human:** Respiratory Irritation (HSDB) Skin Delipidization (HSDB) **Mice:** CNS Depression (HSDB)	Cytochrome P450 Alkaline Phosphatase (ALP) Lipid Peroxidation
Propene	C_3H_6		**Dog:** Cardiac Sensitizer (HSDB) **Human:** CNS Depression (HSDB) Eye/Skin Irritation (HSDB) Resp. Irritation (HSDB) **Mice:** Liver Degeneration (HSDB) **Rat:** Nasal Cavity Lesions (HSDB)	Hydroxylate Reaction
Salicylaldehyde	$C_7H_6O_2$		**Human:** Endocrine Disruption (Sidhu et al. (2005)) **Rat:** Skin Irritant (HSDB)	Enzyme Inhibition EDC

Tetrahydrofuran	C_4H_8O		**Human:** GI Damage (HSDB) Resp. Irritation (HSDB) **CNS** Depression (HSDB) Skin Irritation (HSDB) Eye Irritant (HSDB) **Unknown Species:** Liver Damage (HSDB) Kidney Damage (HSDB) Hypotension (HSDB)	Oxidative Stress Mitogenic Cytochrome P450 Hepatotoxic Aldehyde Oxidation
Triphenylene	$C_{18}H_{12}$		No known disease associations found in literature, all associations are considered suspected disease associaitons **Unknown Species:** Eye Damage (ECHA) Skin Irritant (Haz-Map)	No known mechanisms in literature
Vinyl Acetate	$C_4H_6O_2$		**Human:** Cardiac Irregularities (HSDB) Skin Irritant (HSDB) **Mice:** **Multi Species:** Oral Cancer (HSDB) **Rats:** Nasal Cancer (HSDB) **Rabbit:** Eye Irritant (HSDB) **Unknown Species:** RP Irritant (HSDB) CNS Depression (HSDB) GI Tract Tumors (HSDB) Uterine Tumors (HSDB)	SCE Chromosome Aberration DNA Cross Links Cell Proliferation Cytotoxic Mitotic Inhibition Genotoxic

Note: Table does not contain all chemical-disease associations, just the chemical-disease associations used to test the Brouch Model
*Chemical Properties obtained from ChemSpider http://www.chemspider.com
** Mechanism transcribed from Chapter 2

d. Diseases

The CTD organizes associated diseases into 27 curated disease categories. The 27 disease categories are known as "ancestors." Under the "ancestors," the CTD has associated "descendant" disease. As an example, the "ancestor" disease of chemically induced disease has a "descendant" disease of poisoning. The disease categories used in the Brouch MATLAB® ANN training were taken from the CTD and given a numerical value—one through 27. Similarly to the CTD, the diseases associated with the testing data were selectively matched to the 27 disease categories. Table 2 displays the 27 disease categories used to train the Brouch MATLAB® ANN. Brouch (2014) paired the chemical-disease associations for the curated training data set to the numerical disease categories found in Table 2.

Uncurated testing chemical-disease association and numerical disease category pairing was accomplished using human and animal, medical surveillance and toxicity studies. The medical surveillance and toxicity study findings listed in Table 1 were paired with the disease categories in Table 2 based on the chemical-disease association's comparative anatomical location. The pairing of the uncurated testing data to numerical disease categories can be viewed in Appendix D.

Table 2: Disease Categories

Disease Type/Name	Numerical Disease Category
Animal Diseases	1
Bacterial Infections	2
Cardiovascular Diseases	3
Congenital, Hereditary, Neonatal Diseases	4
Digestive System Diseases	5
Environmental Diseases	6
Endocrine System Diseases	7
Eye Diseases	8
Female Urogenital Disease	9
Hemic and Lymphatic Diseases	10
Immune System Diseases	11
Male Urogenital Diseases	12
Mental Disorders	13
Musculoskeletal Diseases	14
Neoplasms	15
Nervous System Diseases	16
Metabolic Diseases	17
Occupational Diseases	18
Otorhinolaryngologic Diseases	19
Parasitic Diseases	20
Pathological Conditions	21
Respiratory Diseases	22
Skin and Connective Tissue Diseases	23
Stomatognathic Diseases	24
Substance-Related Disorders	25
Virus Diseases	26
Wounds and Injuries	27

e. Remaining Input and Output Data

Before the simulations could be run in MATLAB®, the chemical data and parameters had to be formatted to fit the requirements of the MATLAB® ANN. The training set of chemicals were added to a column in Microsoft® Excel®, followed by each chemical's molecular weight, number of hydrogen bond donors, and number of hydrogen bond acceptors in the corresponding columns. In order to reduce the time that the ANN needed to read the Excel® file, the data was copied from Excel® and pasted into a MATLAB® ".m" file. The ".m" file listed the input data into a single line with an alphabetic code separating each input parameter. Table 3 displays the input and output data for the curated chemical Acetone along with the ".m" file alphabetical code used to identify the parameter's separation. The output data for the testing set were then entered into the corresponding Excel® file columns. The output data consisted of the testing constituent's corresponding disease categories taken from the curated data in the CTD. The output data was then copied from Excel® and pasted into the same MATLAB® ".m" file as the input data.

Table 3: Curated Input and Output Data

Curated Training Data				
	Input			Output
Chemical Name	**Molecular Weight** (UU)	**Hydrogen Acceptors** (VV)	**Hydrogen Donors** (WW)	**Disease Category** (ZZ)
Acetone	58.08	1	0	9
Acetone	58.08	1	0	12
Acetone	58.08	1	0	15
Acetone	58.08	1	0	16
Acetone	58.08	1	0	17
Acetone	58.08	1	0	25

The number of CTD curated chemical-disease associations determined the number of rows used for each training chemical. Using the CTDs curated chemical-disease associations for acetone; the chemical-disease associations were categorized based upon the CTDs 27 disease categories. As seen in Table 3, Acetone has six chemical-disease associations. Since acetone has six chemical-disease associations, there are six rows containing acetone's molecular weight, number of hydrogen bond acceptors, number of hydrogen donor donors, and the corresponding chemical-disease association numerical value. The same process was used for the testing set. However, the testing data set disease associations were not used as outputs, since the MATLAB® ANN was set to predict the associated diseases as an output. An example of the data used for the testing data set can be viewed in Table 4 for Triphenylene. A copy of the complete input and output tables for the training and testing data sets can be found in Appendix A and Appendix B.

Table 4: Uncurated Input and Output Data

Uncurated Testing Data							
	Input			Output			
Chemical (Synonym) (CAS #)	**MW** (CC)	**Hydrogen Bond**		**ANN**		**Known**	
		Acceptor (DD)	**Donor** (EE)	**Derived Disease** (GG)	**Rounded Disease** (HH)	**Disease Category** (II)	**Disease Name**
Triphenylene (Isochrysene) (217-59-4)	228.29	0	0	4.5	5	8	Eye Disease
Triphenylene (Isochrysene) (217-59-4)	228.29	0	0	6	6	23	Skin Disease

2. Artificial Neural Network

ANNs are grouped into two categories: feed-forward and regression. The Brouch MATLAB® ANN is a feed-forward network. Feed-forward networks are organized into layers with connections that flow only in one direction from layer to layer (Jain et al., 1996). The specific type of feed-forward ANN used in the Brouch MATLAB® ANN is a Multilayer Preceptor (MLP). MLPs have three layers: an input layer, a hidden layer, and an output layer (Shamisi et al., 2011). MLPs may have one or more hidden layers which allow the ANN to learn complex non-linear functions (Lee and Lucas, 2014).

In order to learn, feed-forward ANNs need to be trained. Training a feed-forward MLP involves back-propagation. Back-propagation calculates the difference between the actual outputs and predicated outputs and is propagated from the output nodes backwards to the nodes in the previous layer. Back-propagation is accomplished to improve weights during training. Per Adrian Shepherd, a good network will classify patterns similar to, but not identical to, patterns in the training set (Shepherd, 1999). In order to establish well-defined training sets, the number of elements in the input, the target, and the output layers must match.

In order to create a network that can generalize the number of training patterns, the training patterns must be compared to the network weights. If the network weight is greater than the number of training patterns, the ANN may become too powerful and begin over-fitting. The number of hidden layers may also affect network generalization. Having too few hidden networks can leave the network unable to

learn. Having too many hidden layers can result in poor generalization. Lastly, the number of training iterations performed can also alter the network. Having too little iteration, the network will not be able to import data from the training set. In contrast, having too many iterations will result in the network over-training.

In 2014, Brouch developed and programmed a MATLAB® ANN. The formulas were determined by the type of training function specified for the network to use during simulation. The weights and biases were placed on the input data, as the network was tested. After testing the known input and output data with the training function formulas, weight, and biases, the network-derived outputs were compared to the actual output formulated from relevant literature and research.

3. Simulations

Four types of simulations were run through the Brouch MATLAB® ANN: the effect of hidden layers on ANN performance; the effect of the TVT ratios on ANN performance; the effect of chemical structure on ANN performance; and constituent-specific ANN-based predictions for uncurated constituent found in burn pit emissions. This section will briefly describe the type of simulation. Simulation results can be seen and are recorded in Chapter 4 of this thesis.

a. The effect of hidden layers on ANN performance

Initial simulations included running all uncurated constituent simultaneously through the ANN model. Upon input into the model, the data were divided into three subsets: training, validation, and testing (TVT). The TVT used for the initial ANN mode simulations consisted of 70% of the training data, 15% of the validation data, and 15% of the testing data. TVT can also be referenced or

written as 70/15/15. The input into the model was processed through a hidden layer, before the output was generated. Hidden layers consist of hidden neurons that are neither in the input layer nor the output layer and can be conceptualized/visualized in Figures 1 and 2. The utilization of additional hidden layers can be used to determine if the additional layers would increase processing power and system flexibility. However, additional hidden layers can add unwanted or unneeded complexity in the training algorithm. If there are too many hidden layers, there may be more equations than there are free variables resulting in the system being over specified, and the ANN becomes incapable of generalization. If the ANN has too few hidden layers, then the lessened amount of hidden neurons can prevent the system from properly fitting the input data and reduces the robustness of the system. The initial model simulations were capped at five hidden layers to optimize ANN output. Later, simulations were capped at two hidden layers to optimize the run time of the model.

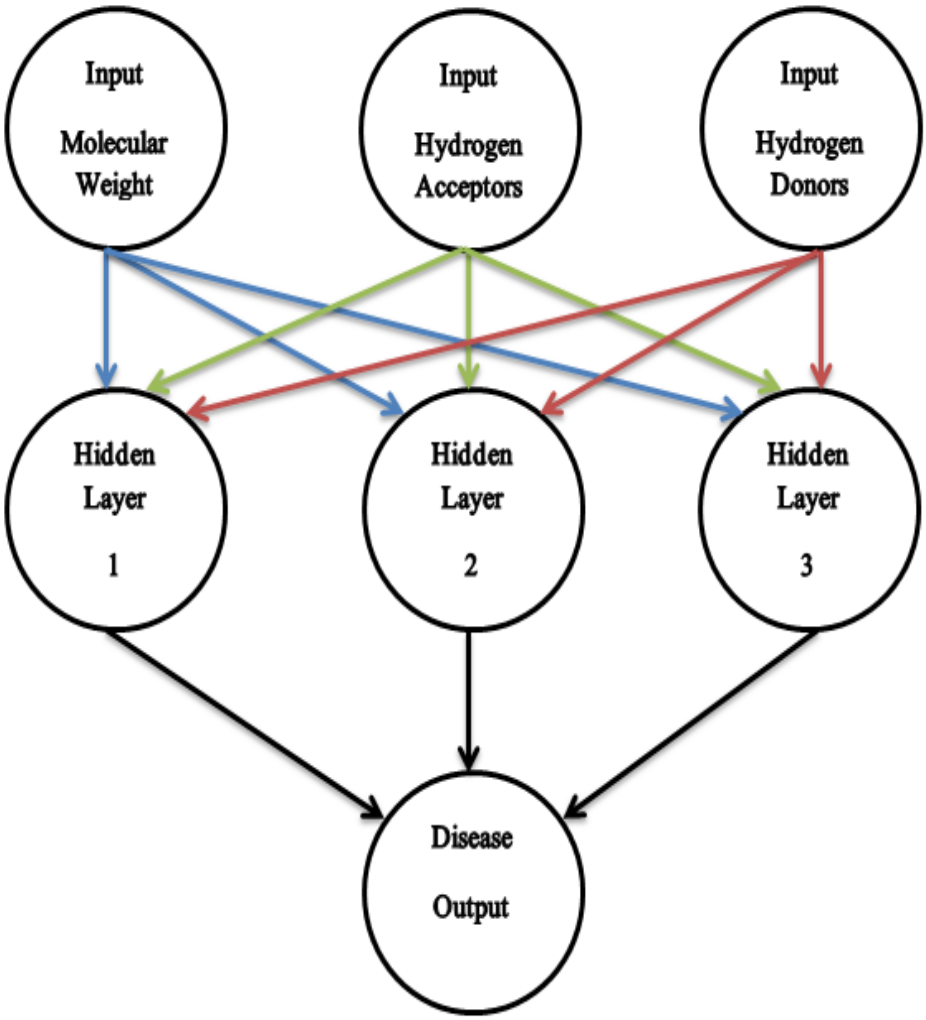


Figure 1: Basic ANN Layout

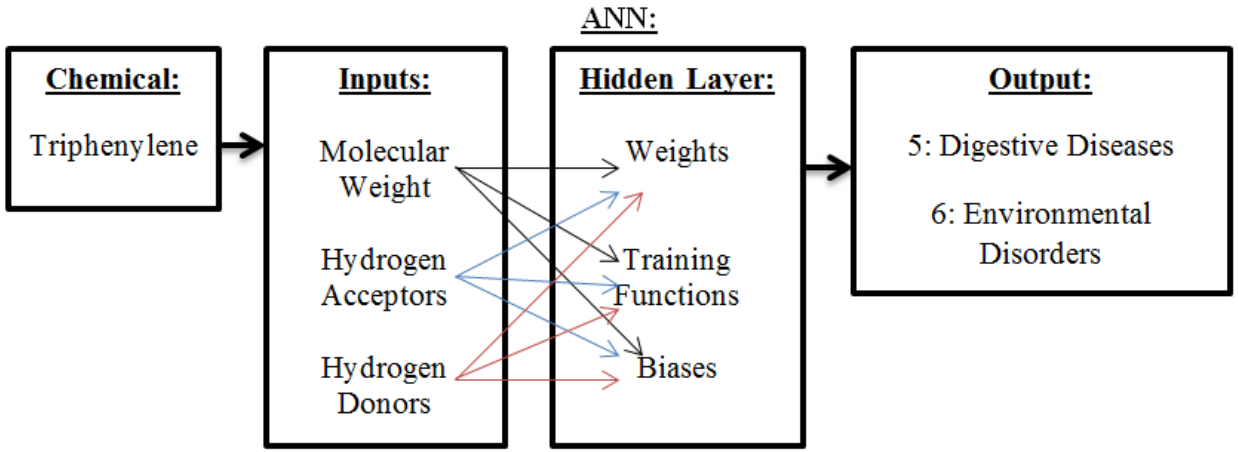


Figure 2: Detailed ANN Layout

b. The effect of the TVT ratios on ANN performance

In order to determine the effect of different training, validation, and testing (TVT) ratios on ANN performance, three TVT ratios were selected and used to run the Brouch ANN: 60/20/20, 70/15/15, and 80/10/10. The training subset represents the percentage of data selected from the "curated training data" that was used to train the model. The first number subset (60, 70, or 80) in the TVT ratio represents the percentage of data used to train the model from the training set. Training occurs by pairing inputs with expected outputs, and is used to compute the gradient, the network weights, and the networks biases. While training the model, the validation error is monitored to prevent the model from over fitting the training data and is derived from the second set of numbers in the TVT ratio (20, 15, or 10). The Brouch ANN uses the uncurated testing data set to evaluate the network performance. The third subset of numbers from the TVT ratio is the testing subset (20, 15, or 10). The testing subset represents the percentage of data selected from the uncurated testing data that was used to generate the models output or results.

c. The effect of chemical structure on ANN performance

Chemicals with similar chemical structure were run through the model at HL =2 and all three TVTs. The three different TVTs used were 60/20/20, 70/15/15, and 80/10/10. The first set of similar structure chemicals (Group 1) included chemicals that contained hydrocarbon chains. The chemicals in Group 1 included: n-heptane, n-octane, and propene. The second set of similar-structure chemicals (Group 2) included benzene ring structures. The chemicals in Group 2

included: 4-ethyltoluene, benzyl chloride, benzanthrone, salicylaldehyde, and triphenylene. Chemical structure similarities can be seen in Table 1.

d. Chemical-specific ANN-based predictions for uncurated constituents found in burn pit emissions

Each uncurated constituent was run through the model individually using HL =2 and all three TVTs: 60/20/20, 70/15/15, and 80/10/10.

4. Analysis and Results

This section will briefly describe how the analysis and results from the MATLAB® ANN were processed. Full analysis and results are recorded in Chapter 4 of this thesis. Once all simulations were complete, the output data was written from the MATLAB® ANN into a Microsoft® Excel® file for analysis. While in Microsoft® Excel®, the ANN output of predicted disease was plotted/graphed against the known associated disease. Once plotted, Excel® was used to calculate the correlation coefficient (R^2). The correlation coefficient indicates the nature and strength of the relationship between the Predicted Disease Category and the Actual Disease Category.

IV. Results and Discussion

Overview

The output generated by the Brouch MATLAB® ANN was recorded in a Microsoft® Excel® scatter plot. The scatter plot was used to graphically represent the relationship between the ANN predicted disease category plotted on the Y-Axis and the known disease category plotted on the X-Axis. The regression equation and the correlation coefficient, or R^2, was calculated using Microsoft® Excel®. The correlation coefficient indicates the nature and strength of the relationship between the predicted disease category and the known disease category. Ultimately, precise agreement between the ANN-predicted disease category and the known disease category was the goal. A correlation coefficient at, or near, the value of one indicates a positive correlation. However, a correlation coefficient at or near the value of one does not indicate that the model can accurately predict chemical disease associations.

The effect of hidden layers on ANN performance

Figure 3 displays the effect of hidden layers on the ANN performance for nine of the ten uncurated constituents found in burn pit emissions. The uncurated constituent, Triphenylene, was excluded from the hidden layer performance modeling due to the absence of known chemical-disease associations. The X-Axis shows the disease category and the Y-Axis shows the predicted disease category that was generated by the ANN. The diagonal line shown on the graph is the reference for precise agreement between the actual and predicted disease categories. The model output was not in agreement with the actual disease categories and was widely scattered, regardless of the number of hidden layers. Numerous incorrect disease associations were observed. When HL =1, there was

a misidentification rate of 84% (37 out of 44), including three misidentifications for chemicals causing nutritional and metabolic diseases and four misidentifications for chemicals causing hemic and lymphatic diseases. Similar observations were made when more than one hidden layer was used. When HL =2, there was a misidentification rate of 73% (32 out of 44), including one misidentifications for chemicals causing stomatognathic diseases and three misidentifications for chemicals causing parasitic diseases. When HL =3, there was a misidentification rate of 93% (41 out of 44), including five misidentifications for chemicals causing bacterial infections and mycoses and for misidentifications for chemicals causing musculoskeletal diseases. When HL =4, there was a misidentification rate of 91% (40 out of 44) including five misidentifications for chemicals causing animal diseases and four misidentifications for chemicals causing environmental disorders. When HL =5, there was a misidentification rate of 89% (39 out of 44), including four misidentifications for chemicals causing immune system diseases and two misidentifications for chemicals causing mental disorders. The regressions of each number of hidden layers exhibited coefficients of determination that were 0.1 or less and may be viewed in Table 5 with each hidden layers regression equation. In light of the high misidentification rates and the low coefficients of determination, these results indicate that the Brouch ANN model lacks the general predictive capability that is needed to screen constituents that are found in burn pit emissions.

To verify consistent model results each HL simulation was run in duplicate. In order to compare the original and duplicate run, the correlation coefficient (R^2) for each run was calculated. Using the R-values from the original and duplicate run, the HL z-score and probability values were calculated (Table 6). A probability value of < 0.05

indicates that the two correlation coefficients for each HL differ significantly. All comparative HL simulations produced probability values in excess of 0.85 indicating that the original and duplicate simulations were statistically similar, demonstrating the internal consistency and reliability of the Brouch ANN.

The effect of additional hidden layers on the performance of the ANN resulted in the unwanted added complexity in the training algorithm. During the hidden layer addition simulations, the network increased training time, which resulted in the network shutting down. Specifically during the HL =4 simulation, the run time lasted longer than 24 hours. Due to this issue, each chemical was run independently at HL =2. The selection of HL =2 maintained simulation run times between 20 and 40 minutes.

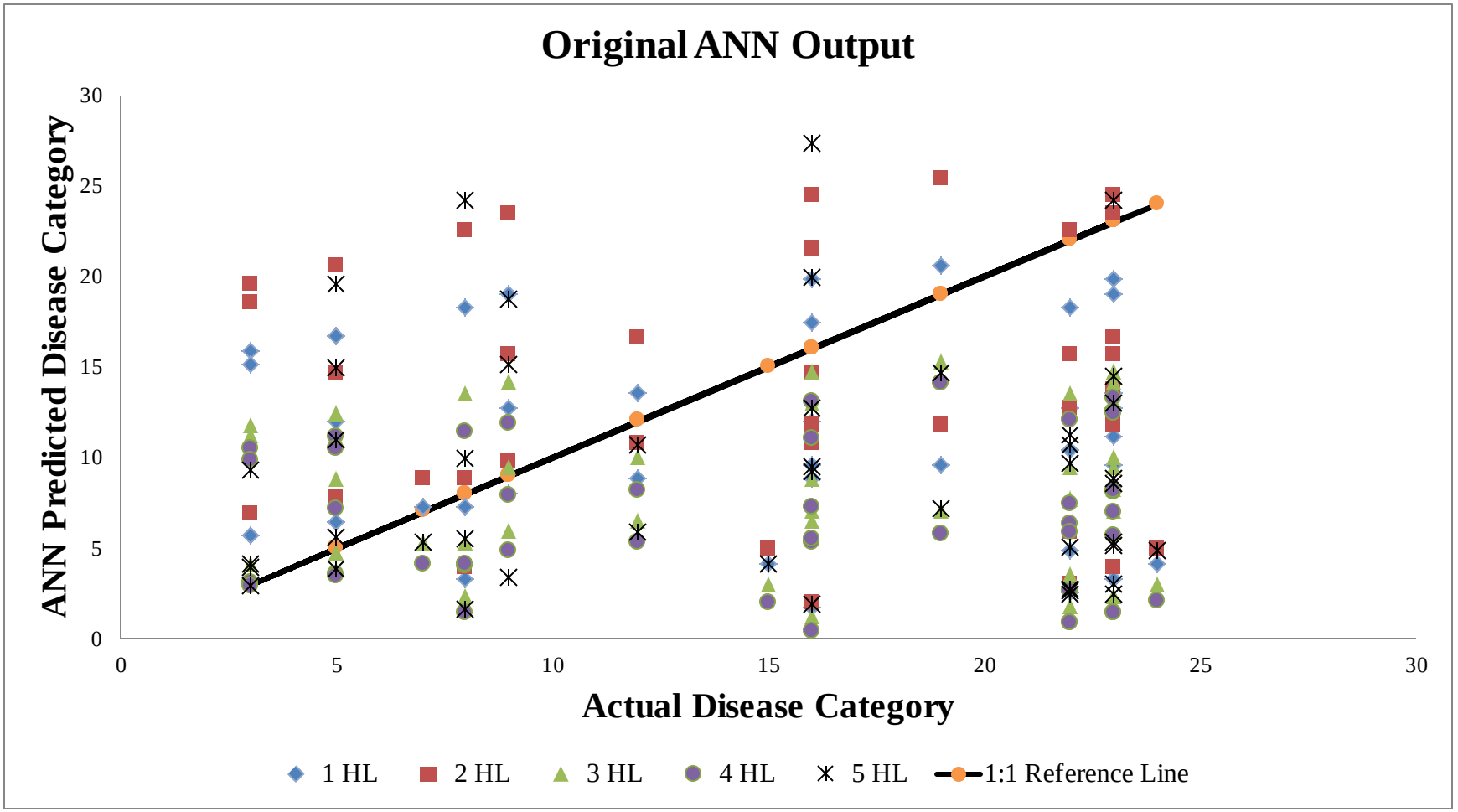


Figure 3: Original ANN Output

Table 5: The effect of the number of hidden layers on ANN performance; Regressions and coefficients of determination

Number of Hidden Layers	Regression Equation	Coefficient of Determination (R^2)
1	y = 0.0016x + 10.787	0.000003
2	y = 0.0016x + 13.229	0.000003
3	y = 0.001x + 7.9777	0.000003
4	y = 0.0021x + 6.5718	0.000005
5	y = -0.0256x + 9.5932	0.0009

Table 6: Determining model reliability using HL simulation comparisons; Regression equations, coefficients of determination, correlation coefficients, z-scores, and probability value

HL	Number of Simulations	Regression Equation	Coefficient of Determination (R^2)	Correlation Coefficient (R)	z-Score	Probability Value
1	1	y = 0.0016x + 10.787	0.000003	0.0017	0.145	0.88
	2	y = 0.0286x + 6.4541	0.1459	-0.0303		
2	1	y = 0.0016x + 13.229	0.000003	0.0017	0.191	0.85
	2	y = 0.3548x + 7.8632	0.1439	-0.0404		
3	1	y = 0.001x + 7.9777	0.000003	0.0017	0.124	0.90
	2	y = 0.2139x + 4.5353	0.1435	-0.0257		
4	1	y = 0.0021x + 6.5718	0.000005	0.0023	0.164	0.87
	2	y = 0.2303x + 4.1325	0.1435	-0.0339		
5	1	y = -0.0256x + 9.5932	0.0009	-0.0294	0.117	0.91
	2	y = 0.3263x + 4.0263	0.0022	-0.0553		

The effect of the TVT ratios on ANN performance

Figures 7 – 11 in Appendix A display the TVT ratio comparisons for nine of the ten uncurated constituents found in burn pit emissions for HL = 1 through HL =5. The uncurated constituent, Triphenylene, was excluded from the TVT performance modeling due to the absence of known chemical-disease associations. The X-Axis displays the disease category and the Y-Axis displays the predicted disease category that was generated by the ANN. The diagonal line shown on the graph is the reference for precise agreement between the actual and predicted disease categories. The model output for all TVT ratio comparisons was not in agreement with the actual disease categories and was widely scattered, regardless of the number of hidden layers. Numerous incorrect disease associations were observed.

The first TVT comparison evaluated the 60/20/20, 70/15/15, and 80/10/10 TVT ratios at HL =1. When the 60/20/20 TVT ratio was used there was a misidentification rate of 100% (44 out of 44). When HL=1 and the 60/20/20 TTV was used the model predicted only negative disease category values. When the 70/15/15 TVT ratio was used there was a misidentification rate of 84% (37 out of 44), including three misidentifications for chemicals causing nutritional and metabolic disorders and two misidentifications for chemicals causing occupational diseases. When the 80/10/10 TVT ratio was used there was a misidentification rate of 84% (37 out of 44), including two misidentifications for chemicals causing parasitic diseases and three misidentifications for chemicals causing immune system diseases. The TVT comparisons for HL =1 can be

viewed in Figure 7 of Appendix A. Similar observations were made when more than one hidden layer was used.

The second TVT comparison evaluated the 60/20/20, 70/15/15, and 80/10/10 TVT ratios at HL =2. When the 60/20/20 TVT ratio was used there was a misidentification rate of 100% (44 out of 44). When HL=2 and the 60/20/20 TTV was used, the model predicted only negative disease category values. When the 70/15/15 TVT ratio was used, there was a misidentification rate of 77% (34 out of 44), including two misidentifications for chemicals causing pathological conditions and two misidentifications for chemicals causing nutritional and metabolic diseases. When the 80/10/10 TVT ratio was used there was a misidentification rate of 75% (33 out of 44), including two misidentifications for chemicals causing substance-related disorders and one misidentifications for chemicals causing virus diseases. The TVT comparisons for HL =2 can be viewed in Figure 8 of Appendix A.

The third TVT comparison evaluated the 60/20/20, 70/15/15, and 80/10/10 TVT ratios at HL =3. When the 60/20/20 TVT ratio was used there was a misidentification rate of 100% (44 out of 44). When HL=3 and the 60/20/20 TTV was used the model predicted only negative disease category values. When the 70/15/15 TVT ratio was used there was a misidentification rate of 86% (38 out of 44), including four misidentifications for chemicals causing musculoskeletal diseases and two misidentifications for chemicals causing neoplasms. When the 80/10/10 TVT ratio was used there was a misidentification rate of 84% (37 out of 44), including five misidentifications for chemicals causing bacterial infections and mycoses and three misidentifications for chemicals causing

mental disorders. The TVT comparisons for HL =3 can be viewed in Figure 9 of Appendix A.

The fourth TVT comparison evaluated the 60/20/20, 70/15/15, and 80/10/10 TVT ratios at HL =4. When the 60/20/20 TVT ratio was used there was a misidentification rate of 96% (42 out of 44), including four misidentifications for chemicals causing bacterial infections and mycoses. When the 70/15/15 TVT ratio was used there was a misidentification rate of 87% (39 out of 44), including four misidentifications for chemicals causing immune diseases and three misidentifications for chemicals causing animal diseases. When the 80/10/10 TVT ratio was used there was a misidentification rate of 86% (38 out of 44), including three misidentifications for chemicals causing hemic and lymphatic diseases and four misidentifications for chemicals causing mental disorders. The TVT comparisons for HL =4 can be viewed in Figure 10 of Appendix A.

The final TVT comparison evaluated the 60/20/20, 70/15/15, and 80/10/10 TVT ratios at HL =5. When the 60/20/20 TVT ratio was used there was a misidentification rate of 98% (43 out of 44). When HL=3 and the 60/20/20 TTV was used the model predicted 29 negative disease category values. When the 70/15/15 TVT ratio was used there was a misidentification rate of 91% (40 out of 44), including four misidentifications for chemicals causing bacterial infections and mycoses and two misidentifications for chemicals causing parasitic diseases. When the 80/10/10 TVT ratio was used there was a misidentification rate of 91% (40 out of 44) including four misidentifications for chemicals causing environmental disorders and one misidentification for chemicals causing substance-related disorders. The TVT comparisons for HL =5 can be viewed in Figure 11 of Appendix A.

The regressions for each TVT ratio exhibited coefficients of determination that were 0.1 or less and are displayed in Table 7. Considering the high misidentification rates and the low coefficients of determination, the TVT comparison results indicate that the TVT ratio does not impact the accuracy of the Brouch ANN. To further demonstrate this point, the bar graph in Figure 4 displays all TVT comparisons for each of the hidden layers tested, where the X-Axis displays the results of each hidden layers TVT comparisons, and the Y-Axis displays the coefficient of determination (R^2).

Table 7: The effect of TVT ratios on ANN performance; Regressions and misidentification ratios

HL	Regression Equation			Misidentification Ratio		
	60/20/20	70/15/15	80/10/10	60/20/20	70/15/15	80/10/10
1	y = 0.0123x + 4.908	y = -0.026 + 11.059	y = -0.0266x + 11.227	44:44	37:44	37:44
2	y = -5E-05x – 0.0086	y = -0.0495x + 13.854	y = -0.0384x + 13.874	44:44	34:44	33:44
3	y = 0.0293x – 9.2427	y = -0.0184x + 8.107	y = -0.028x + 8.2673	44:44	38:44	37:44
4	y = -0.0521x + 2.8549	y = -0.042x + 6.5502	y = -0.021x + 7.9868	42:44	39:44	38:44
5	y = -0.021x – 0.8091	y = -0.0485x + 9.8617	y = -0.0468x + 10.02	43:44	40:44	40:44

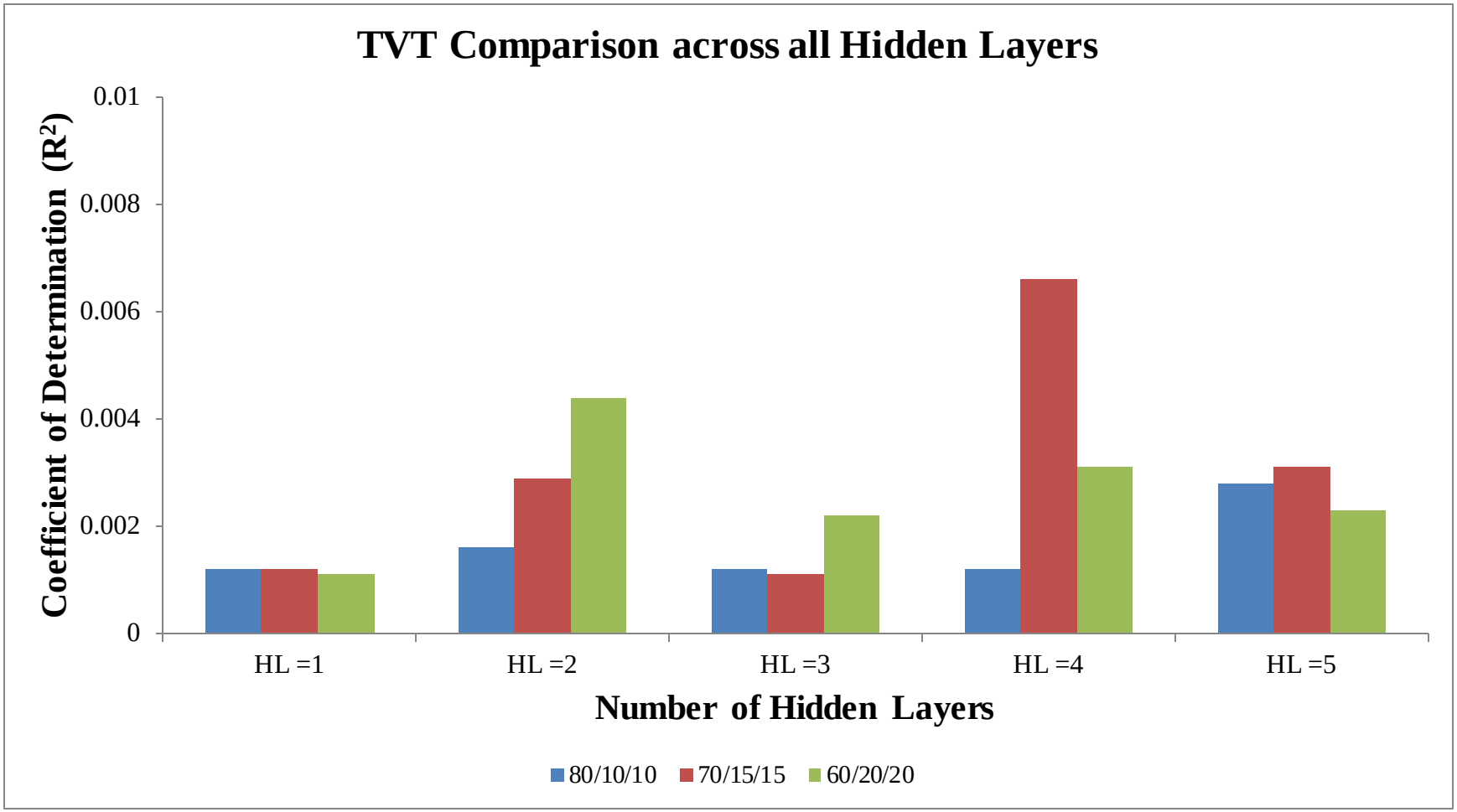


Figure 4: TVT Comparison across all HLs

The effect of chemical structure on ANN performance

Figures 5 and 6 show the effects of chemical structure on the performance of the ANN for eight of the ten uncurated constituents found in burn pit emissions at HL =2. The eight uncurated constituents tested contained either benzene rings or hydrocarbon chains. The two excluded uncurated constituents; tetrahydrofuran and vinyl acetate, were excluded from the similar structure performance modeling due to their chemical structure. The X-Axis displays the known disease category and the Y-Axis displays the predicted disease category that was generated by the ANN. The diagonal line shown on the graph is the reference for precise agreement between the actual and predicted disease categories. The model output was not in agreement with the actual disease categories and was widely scattered, regardless of the TVT ratio. Numerous incorrect disease associations were observed.

Group 1 – Chemical Structures containing Hydrocarbon Chains

The chemicals in Group 1 contained hydrocarbon chains and included the following chemicals: n-heptane, n-octane, and propene. Group 1 chemicals were evaluated at the 60/20/20, 70/15/15, and 80/10/10 TVT ratios at HL =2. Comparing the hydrocarbon chained chemicals from the Group 1 data set to the Curated Training Data set; the Curated Training Data set did not contain chemicals that were structurally related to the Group 1 data set.

When Group 1 chemicals were modeled at the 60/20/20 TVT ratio there was a misidentification rate of 100% (15 out of 15), including four misidentifications for chemicals causing bacterial infections and mycoses and

three misidentifications for chemicals causing endocrine system diseases. When the Group 1 chemicals were modeled at the 70/15/15 TVT ratio there was a misidentification rate of 73% (11 out of 15), including two misidentifications for chemicals causing female urogenital diseases and pregnancy complications. When compared to the original ANN output at the 70/15/15 TVT ratio for HL =2 simulation captured in Figure 3, the two simulations share similar misidentification ratios. The chemicals in Group 1 at the 70/15/15 for HL =2 had a misidentification rate of 73% (11 out of 15), while the original ANN output at the 70/15/15 TVT for HL =2 had a misidentification rate of 73% (32 out of 44). When Group 1 chemicals were modeled at the 80/10/10 TVT ratio there was a misidentification rate of 80% (12 out of 15), including two misidentifications for chemicals causing environmental disorders and two misidentifications for chemicals causing hemic and lymphatic diseases. The Group 1 TVT comparisons at HL =2 can be viewed in Figure 5.

Group 2 – Chemical Structures containing Benzene Rings

The chemicals in Group 2 contained benzene rings and included the following chemicals: 4-ethyltoluene, benzyl chloride, benzanthrone, salicylaldehyde, and triphenylene. Group 2 chemicals were evaluated at the 60/20/20, 70/15/15, and 80/10/10 TVT ratios at HL =2. Comparing the benzene ring containing chemicals from the Group 2 data set to the curated training data set, 55% (41 out of 75), chemicals from the Curated Training Data contained benzene ring structures.

When group 2 chemicals were modeled at the 60/20/20 TVT ratio there was a misidentification rate of 86% (12 out of 14), including two misidentifications for chemicals causing congenital, hereditary, neonatal diseases and abnormalities and three misidentifications for chemicals causing immune system diseases. When Group 2 chemicals were modeled at the 70/15/15 TVT ratio there was a misidentification rate of 100% (14 out of 14). The 70/15/15 TVT ratio predicted negative disease category values less than one. When compared to the original ANN output at the 70/15/15 TVT ratio for HL =2 simulation captured in Figure 3, the two simulations did not share similar misidentification ratios. The chemicals in Group 2 at the 70/15/15 for HL =2 had a misidentification rate of 100% (14 out of 14), while the original ANN output at the 70/15/15 TVT for HL =2 had a misidentification rate of 73% (32 out of 44). When Group 2 chemicals were modeled at the 80/10/10 TVT ratio there was a misidentification rate of 86% (12 out of 14) including four misidentifications for chemicals causing immune system diseases and three misidentifications for chemicals causing eye diseases. The Group 2 TVT comparisons at HL =2 can be viewed in Figure 6.

The regression equations and coefficients of determination for Groups 1 and 2 are displayed in Table 8. The regressions for each groups TVT ratio displayed coefficients of determination that were 0.1 or less. Considering the high misidentification rates and the low coefficients of determination, the similar chemical structure comparisons indicate that a chemical structure similarity does not impact the accuracy of the Brouch ANN. At this time, it is recommended that

the curated training data set be expanded to include chemicals that contain hydrocarbon chains.

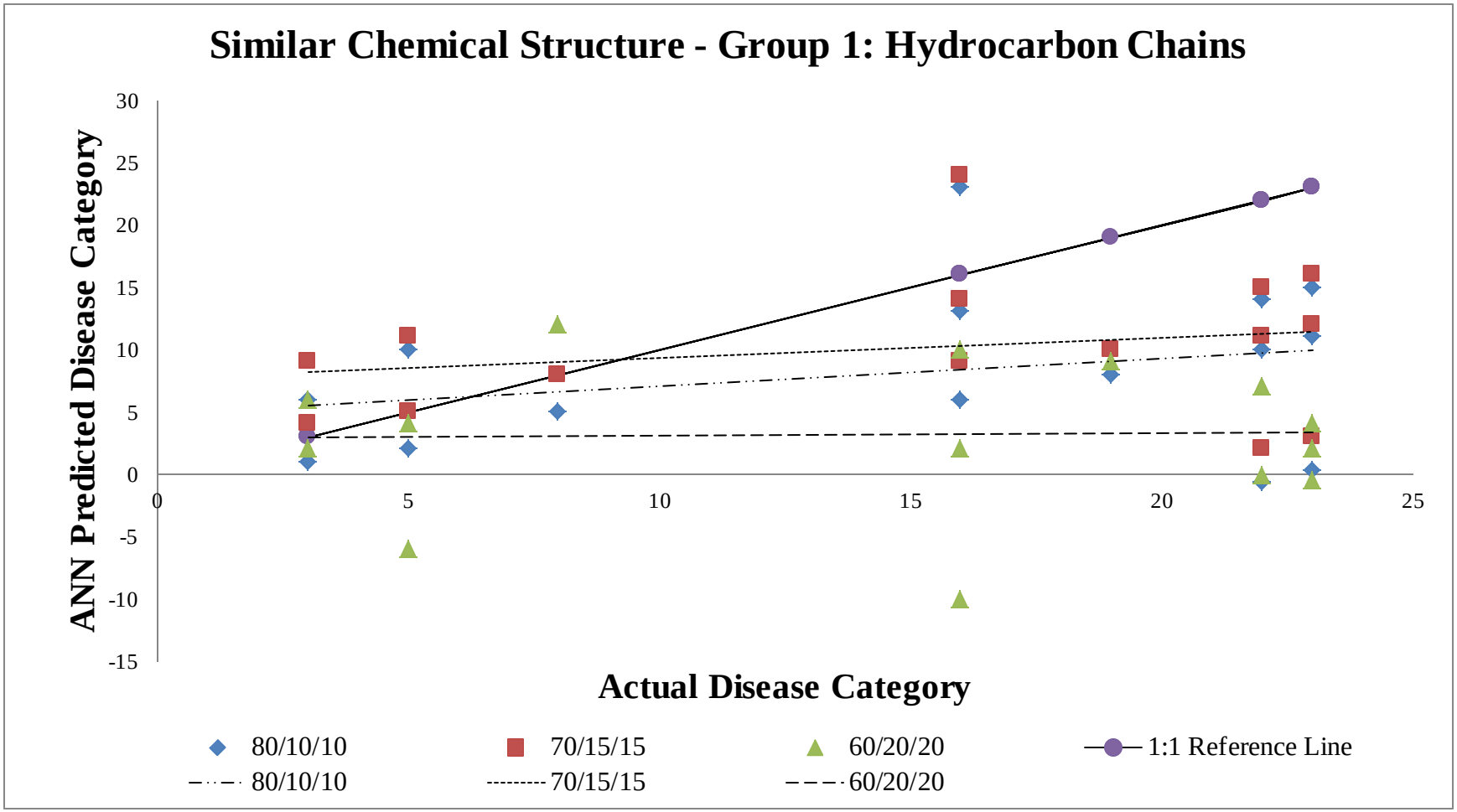


Figure 5: Similar Chemical Structure - Group 1: Hydrocarbon Chains

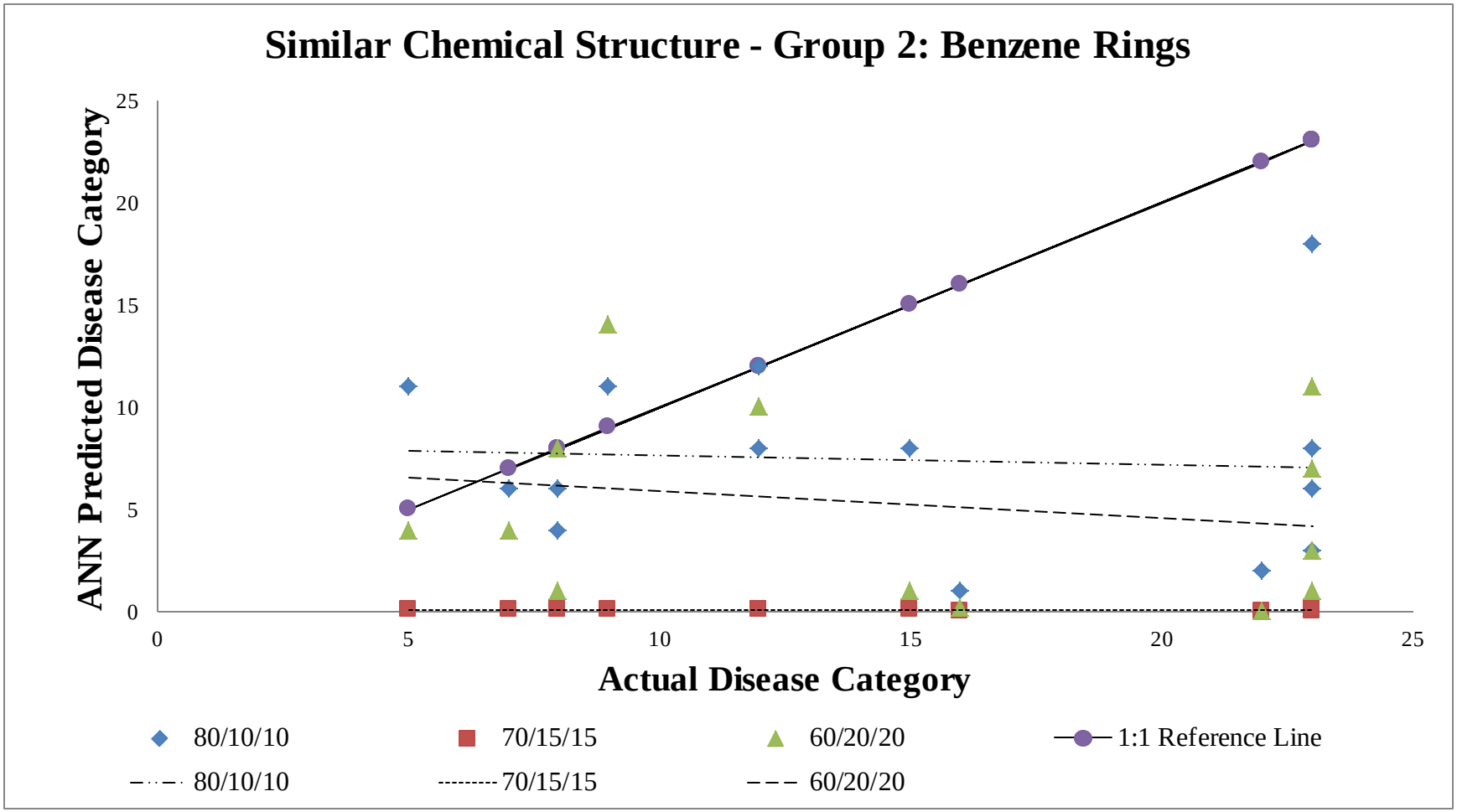


Figure 6: Similar Chemical Structure - Group 2: Benzene Rings

Table 8: The effect of Chemical Structure on ANN performance; Regressions and misidentification ratios

Group	Regression Equation			Misidentification Ratio		
	60/20/20	70/15/15	80/10/10	60/20/20	70/15/15	80/10/10
1	y = 0.018x + 2.9561	y = 0.1643x + 7.6949	y = 0.2238x + 4.8685	15:15	11:15	12:15
2	y = 0.1313x + 7.2329	y = -0.0008x + 0.1094	y = 0.0472x + 8.1226	12:14	14:14	12:14

Chemical-specific ANN-based predictions for uncurated constituents found in burn pit emissions

Misidentification Ratios

Figures 12 through 21 in Appendix B display the effects of individual constituents on the ANN performance for the ten uncurated constituents found in burn pit emissions at HL =2. All uncurated constituents were evaluated at the 60/20/20, 70/15/15, and 80/10/10 TVT ratios. The X-Axis displays the known disease category and the Y-Axis displays the predicted disease category that was generated by the ANN. The diagonal line shown on the graph is the reference for precise agreement between the actual and predicted disease categories. The model output was not in agreement with the actual disease categories and was widely scattered, regardless of the TVT ratio. Numerous incorrect disease associations were observed. Table 9 displays the uncurated constituents, disease associations, mechanism of disease, and the misidentification ratios for the individual uncurated constituent simulations.

4-Ethyltoluene

When the 60/20/20 TVT ratio was used there was a misidentification rate of 100% (3 out of 3), to include the misidentification of bacterial infections, endocrine system, and metabolic diseases. When the 70/15/15 TVT ratio was used there was a misidentification rate of 100% (3 out of 3), to include the misidentification of female urogenital diseases, and neoplasms. When the 80/10/10 TVT ratio was used there was a misidentification rate of 100% (3 out of 3), to include the

misidentification of digestive system, endocrine system, and eye diseases. The TVT comparisons for 4-ethyltoluene can be viewed in Figure 12 of Appendix B.

Benzanthrone

When the 60/20/20 TVT ratio was used there was a misidentification rate of 100% (4 out of 4), to include the misidentification of immune system diseases and three model predictions for negative disease category values. When the 70/15/15 TVT ratio was used there was a misidentification rate of 75% (3 out of 4), to include the misidentification of digestive system disease, and environmental disorders. When the 80/10/10 TVT ratio was used there was a misidentification rate of 100% (4 out of 4), to include the misidentification of eye, immune system, and musculoskeletal diseases. The TVT comparisons for benzanthrone can be viewed in Figure 13 of Appendix B.

Benzyl Chloride

When the 60/20/20 TVT ratio was used there was a misidentification rate of 100% (3 out of 3), to include the misidentification of immune system diseases and two model predictions for negative disease category values. When the 70/15/15 TVT ratio was used there was a misidentification rate of 67% (2 out of 3), to include the misidentification of environmental disorders, and digestive system diseases. When the 80/10/10 TVT ratio was used there was a misidentification rate of 67% (2 out of 3), to include the misidentification of immune system and

musculoskeletal diseases. The TVT comparisons for benzyl chloride can be viewed in Figure 14 of Appendix B.

n-Heptane

When the 60/20/20 TVT ratio was used there was a misidentification rate of 80% (4 out of 5), to include the misidentification of neoplasms, immune system, and musculoskeletal diseases. When the 70/15/15 TVT ratio was used there was a misidentification rate of 80% (4 out of 5), to include the misidentification of environmental, endocrine system, and eye diseases. When the 80/10/10 TVT ratio was used there was a misidentification rate of 80% (4 out of 5), to include the misidentification of cardiovascular, congenital, hereditary, neonatal, environmental, hemic, and lymphatic diseases. The TVT comparisons for n-heptane can be viewed in Figure 15 of Appendix B.

n-Octane

When the 60/20/20 TVT ratio was used there was a misidentification rate of 100% (3 out of 3), to include the misidentification of congenital, hereditary, neonatal, animal, and cardiovascular diseases. When the 70/15/15 TVT ratio was used there was a misidentification rate of 100% (3 out of 3), to include the misidentification of animal diseases, bacterial infections, and one model prediction for negative disease category values. When the 80/10/10 TVT ratio was used there was a misidentification rate of 100% (3 out of 3), to include the misidentification

of neoplasms, endocrine system, and immune system diseases. The TVT comparisons for n-octane can be viewed in Figure 16 of Appendix B.

Propene

When the 60/20/20 TVT ratio was used there was a misidentification rate of 71% (5 out of 7), to include the misidentification of congenital, hereditary, neonatal, endocrine system, and environmental diseases. The 60/20/20 TVT ratio also produced the misidentification of bacterial infections, and one model prediction for negative disease category values. When the 70/15/15 TVT ratio was used there was a misidentification rate of 71% (5 out of 7), to include the misidentification of animal, immune system, environmental, female urogenital, hemic, and lymphatic diseases. When the 80/10/10 TVT ratio was used there was a misidentification rate of 71% (5 out of 7), to include the misidentification of animal, endocrine system, female urogenital and musculoskeletal diseases. The TVT comparisons for propene can be viewed in Figure 17 of Appendix B.

Salicylaldehyde

When the 60/20/20 TVT ratio was used there was a misidentification rate of 100% (2 out of 2), to include the misidentification of environmental diseases, and one model prediction for negative disease category values. When the 70/15/15 TVT ratio was used there was a misidentification rate of 100% (2 out of 2), to include the misidentification of immune system diseases. When the 80/10/10 TVT ratio was used there

was a misidentification rate of 100% (2 out of 2), to include the misidentification of environmental diseases. The TVT comparisons for salicylaldehyde can be viewed in Figure 18 of Appendix B.

Tetrahydrofuran

When the 60/20/20 TVT ratio was used there was a misidentification rate of 75% (6 out of 8), to include the misidentification of substance-related disorders, endocrine system, and metabolic diseases. When the 70/15/15 TVT ratio was used there was a misidentification rate of 50% (4 out of 8), to include the misidentification of neoplasms, substance-related disorders, and metabolic diseases. When the 80/10/10 TVT ratio was used there was a misidentification rate of 63% (5 out of 8), to include the misidentification of bacterial infections, endocrine system, eye, and musculoskeletal diseases. The TVT comparisons for tetrahydrofuran can be viewed in Figure 19 of Appendix B.

Triphenylene

When the 60/20/20 TVT ratio was used there was a misidentification rate of 100% (2 out of 2), to include the misidentification of immune system diseases, and one model prediction for negative disease category values. When the 70/15/15 TVT ratio was used there was a misidentification rate of 100% (2 out of 2), to include the misidentification of digestive system, and environmental diseases. When the 80/10/10 TVT ratio was used there was a misidentification rate of 100% (2 out of 2), to include the misidentification of immune system, and eye diseases. The

TVT comparisons for triphenylene can be viewed in Figure 20 of Appendix B.

Vinyl Acetate

When the 60/20/20 TVT ratio was used there was a misidentification rate of 89% (8 out of 9), to include the misidentification of bacterial infections, neoplasms, pathological conditions, congenital diseases, hereditary diseases, neonatal disease, and one model prediction for negative disease category values. When the 70/15/15 TVT ratio was used there was a misidentification rate of 89% (8 out of 9), to include the misidentification of bacterial infections, environmental diseases, one half-value model prediction, and three model predictions for negative disease category values. When the 80/10/10 TVT ratio was used there was a misidentification rate of 89% (8 out of 9), to include the misidentification of animal, endocrine system, immune system, and metabolic diseases, along with the prediction of two negative disease category values. The TVT comparisons for vinyl acetate can be viewed in Figure 21 of Appendix B.

Table 9: Constituent specific ANN-based predictions for uncurated chemicals found in burn pit emissions: Disease causation, mechanisms, and misidentification ratios

Constituents	Disease Causation	Mechanisms Of Action	Misidentification Ratio		
			60/20/20	70/15/15	80/10/10
4-ethyltoluene	16, 22, 23	Sister Chromatid Exchange (SCE), Gamma-Glutamyl Transferase (GGT), and Lactic Acid Dehydrogenase (LDH)	3:3	3:3	3:3
Benzanthrone	5, 9, 12, 23	Lipid Peroxidation, Oxidative Stress, Endocrine Disruption (EDC), Immunotoxin	4:4	3:4	4:4
Benzyl Chloride	5, 22, 23	Alkylating Agent, Lipid Peroxidation, Chromatid Aberration	3:3	2:3	2:3
n-Heptane	3, 5, 16, 22, 23	Cytochrome P450, Oxidation	4:5	4:5	4:5
n-Octane	16, 22, 23	Cytochrome P450, Alkaline Phosphatase (ALP), Lipid Peroxidation	3:3	3:3	3:3
Propene	3, 5, 8, 16, 19, 22, 23	Hydroxylate Reaction	5:7	5:7	5:7
Salicylaldehyde	7, 23	Enzyme Inhibition, EDC	2:2	2:2	2:2
Tetrahydrofuran	3, 5, 8, 9, 12, 16, 22, 23	Oxidative Stress, Mitogenic, Cytochrome P450, Hepatotoxic Aldehyde	6:9	4:9	5:9
Triphenylene	8, 23	No known mechanisms in literature	2:2	2:2	2:2
Vinyl Acetate	3, 5, 8, 9, 16, 19, 22, 23, 24	SCE, Chromosome Aberration, DNA Cross Links, Cell Proliferation, Cytotoxic, Mitotic Inhibition, Genotoxic	8:9	8:9	8:9

Data Gaps

Chemical-Disease-Mechanism Associations

The chemical-disease associations used to train the Brouch ANN model were selected due to their known associations described in the CTD. However, the chemical-disease associations selected to test the Brouch ANN model were derived from suspected disease-associations due to the limited nature of available literature and research. The high misidentification ratios indicate that more data is needed for ANN training with uncurated constituents.

Chemical Mixtures

Due to the limited data on chemical mixtures and their side effects, the Brouch ANN was not trained to predict disease for chemical mixtures. As it stands, the ten uncurated burn pit constituents used to test the Brouch ANN enter the body in the form of a mixture. However, the burn pit constituents were modeled as individual chemical components due to the trained ability of the model. With the understanding that each chemical maintains its own toxic effect inside the body, the chemical combination present in mixtures causes the mechanisms and health predictions to become increasingly complicated. The combination of mechanisms and toxic effects may result in the additive, synergistic, or potentiating effects. Additive effects occur when chemicals have similar toxic effects, produce a combined effect that is equal to the sum of the chemicals separate

effects. Synergistic effects occur when the chemical combination produces a health effect that is greater than the sum of the individual chemical effects. Potentiating effects occur when an effect of one chemical is increased by the exposure to a second chemical. Data and chemical testing is limited on these types of chemical interactions resulting in the inability to model the health effects of chemical mixtures. Expanding the training data set to include chemical mixtures may help overcome this limitation.

Further Testing

Further testing is recommended for the burn pit constituents with R^2 values greater than 0.8. The constituents with an R^2 value greater than 0.8 include: 4-ethyltoluene, benzanthrone, n-heptane, n-octane, salicylaldehyde, and triphenylene. The six constituents identified above can be viewed in Table 10. These six constituents should be prioritized to help further develop predictive ANN models for human health force support.

The American Chemistry Council's (ACC's) 2011 Chemical Prioritization Screening Approach (ACC, 2011) was adopted to prioritize the six uncurated burn pit emission constituents. In order to rank and assign a priority to a constituent, the ACC recommends identifying the following: the constituent's toxicity, carcinogenicity, endocrine disruptor status, and available data. These items were used to rank the health hazards for the six uncurated burn pit constituents and can be seen in

Table 10. The persistency and bioaccumulation variable for each constituent were also used as indicators of exposure. The persistency and bioaccumulation criteria from the ACC approach, recommended distinguishing between persistent and non-persistent chemicals using the following criteria: volatile chemicals are maintaining a vapor pressure > 1000 Pa. Persistent versus non-persistent chemicals were differentiated using a chemical half-life in air.

A constituent was not considered persistent if the air half-life was < 2 days. The air half-life was derived from Scheringer et al. (2006). Scheringer et al. (2006) purposed the air half-life of <2 days as a screening criterion for chemical degradability in air. The selection of an air half-life of 2 days was based, in part, on the distance traveled by a chemical in the air. Scheringer et al. (2006) used the AOPWIN^{TM} estimation software to estimate the air half-life of organic chemicals that do not have measured rate constants. AOPWIN^{TM} estimates chemical atmospheric oxidation potential, and is currently used for the European Union's registration, evaluation and authorization of chemicals (REACH) program. The air half-life for the six uncurated constituents was derived using the AOPWIN^{TM} model obtained from the Environmental Protection Agency's EPIsuite^{TM} available on the ChemSpider website. The biodegradability for the six uncurated constituents was derived using the BIOWIN^{TM} model obtained from the Environmental Protection Agency's EPIsuite^{TM} available on the ChemSpider website. BIOWIN^{TM} estimates the chemical

biodegradability. The ACC in the 2011 Chemical Prioritization Screening Approach (ACC, 2011) suggests the use of the AOPWIN™ and BIOWIN™ models.

To assess for bio-concentration, the bi- concentration factor (BCF), from the ACD/Labs Percepta Platform—Software Modules was used. The ACD/Labs Percepta Platform—Software Modules predict physicochemical, ADME, and toxicity properties from chemical structure. ACD/BCF was accessed through the ChemSpider website. In order to classify constituents as a bio-concentrator, a chemical must produce a bio-concentration factor (BCF) > 5000. The biodegradable, half-life, persistency, and bio-concentration factors for the six uncurated constituents can be viewed in Table 10.

Based on the above recommendations, the six uncurated constituents were grouped with regard to persistence and bio-concentration according to the factors in Table 11. As demonstrated in Table 11, each constituent was assigned a numerical score based upon the constituent's overall ranking. The uncurated constituent persistency and bio-concentration scores can be seen in Table 14. The six uncurated chemicals were then grouped with regard to human health hazards and the factors found in Table 12. As demonstrated in Table 12, each constituent was assigned a numerical score based upon the constituent's ranking. The human health hazards scores can be seen in Table 14. The persistence and bio-concentration scores and the human health hazards score were summed

to derive an overall final score or value. These values were then separated into categories from low to high and given a priority level. The final score and priority level for the six uncurated burn pit constituents can be seen in Table 14.

According to the priority level identified in Table 14, salicylaldehyde and triphenylene should be the first of the six constituents to require additional sampling and testing. The second group to obtain additional sampling and evaluation are benzanthrone and n-octane. The last group to attain further assessment is 4-ethyltoluene and n-heptane.

Table 10: Uncurated constituent human health hazard classification criteria

Constituent	Adapted from ACC Human Health Hazard Classification Criteria									R^2		
	Toxicity		Carcinogen	Insufficient Chemical Information	Biodegradable	½ life (days)	Bio-concentration Factor (BCF)	Persistency VP (Pa)	EDC	60/20/20	70/15/15	80/10/10
	Acute	Chronic										
4-ethyltoluene	---	---	---	Yes	No	4.9	No (451)	No (0.67)	No	0.10	0.85	0.96
Benzanthrone	---	Yes	---	Yes	No	-	No (2350)	No (0.67)	Yes	0.75*	0.84	0.88
n-Heptane	Yes	Yes	No	Not tested for Carcinogenicity	Yes	5.5	No (1502)	No (0.00003)	No	0.94	0.98	0.70
n-Octane	Yes	Yes	No	Yes	Yes	6.4	No (3825)	Yes (6133)	No	0.02*	1.00*	0.96
Salicylaldehyde	---	---	---	Yes	Yes	-	No (10 – 11)	Yes (1880)	Yes	1.00*	1.00	1.00
Triphenylene	---	---	Possible Mutagen	Yes	No	344	Yes (13349)	No (79)	No	1.00*	1.00	1.00

Note: Properties obtained from ChemSpider http://www.chemspider.com
(---) Denotes no information available
(-) Denotes incompatibility with BIOWIN Model
(*) Denotes negative linearity

Table 11: Persistence and bio-concentration exposure ranking

Persistence and Bio-concentration	Ranking	Persistence and Bio-concentration Score
Not Persistent and No Bio-concentration	Low	1
Persistent and No Bio-concentration OR Bio-concentration and Not Persistent	Medium	3
Persistent and Bio-concentration	High	5

Table 12: Human health hazard ranking

Human Health	Ranking	Health Ranking Score
Non-Carcinogen with Acute Toxicity	Medium	2
Non-Carcinogen with Chronic Toxicity	Medium	2
Non-Carcinogen w/ Acute & Chronic Toxicity	Medium -High	3
Insufficient Data to Classify	High	4
Carcinogen	No ranking add 1 to final score	1
EDC Chemical	No ranking add 1 to final score	1

Table 13: Integration of exposure rankings

Combined Score	Exposure Ranking	Priority
2 – 4	Low	Low
5 – 7	Medium	Medium
8 – 10	High	High

Table 14: Uncurated constituent proposed prioritization

Constituent	Adapted from ACC Human Health Hazard Classification Criteria					
	Persistence and Bio-concentration	Human Health	Carcinogen	EDC	Final Score	Priority
4-ethyltoluene	1	4	-	-	5	Low
Benzanthrone	1	4	-	1	6	Med
n-Heptane	1	3	-	-	4	Low
n-Octane	3	3	-	-	6	Med
Salicylaldehyde	3	4	-	1	8	High
Triphenylene	3	4	1	-	8	High

V. Conclusions and Recommendations

Conclusions and Recommendations

1. Misidentification rates of 73% or greater were observed for ANN simulations when the hidden layer size varied between 1 and 5. The ANN model, as currently constructed and trained, does not have the predictive capability needed to screen constituents associated with burn pit emissions. The number of hidden layers had little effect on the performance of the model. Additional data is needed to train the model.
2. Misidentification rates of 75% or greater were observed for ANN simulations when the TVT ratios ranged from 60/20/20 to 80/10/10 for the nine uncurated test constituents. As currently constructed and trained, the ANN model does not have the predictive capability needed to screen burn pit emission constituents. The TVT ratios had little effect on the performance of the model, and are likely due to the need for additional training data.
3. ANN-based screening of similar structured constituents containing benzene rings and hydrocarbon chains produced misidentification rates of 73% or greater, and R^2 values of 0.0762 and lower. The misidentification rates for similar structure constituents were lower than those misidentification rates observed for the original test set, while the R^2 values for similar structure constituents were higher than the R^2 values observed for the original test set.
4. Further testing is recommended for several constituents with R^2 values greater than 0.8. The constituents with an R^2 value greater than 0.8 include 4-ethyltoluene, benzanthrone, n-heptane, n-octane, salicylaldehyde, and

triphenylene. These constituents have been prioritized to help further develop predictive ANN models for human health force support. Salicylaldehyde and triphenylene should be the first constituents to acquire additional sampling and testing, followed by benzanthrone and n-octane, and 4-ethyltoluene and n-heptane.

5. Evaluation of alternative disease categorization approaches is needed to determine the impact of disease categorization of the model's performance.
6. Reclassifying the training and testing data from categorical variable to continuous variables is needed to validate the chemical-diseases mapping provided by the Brouch model.

Appendix A

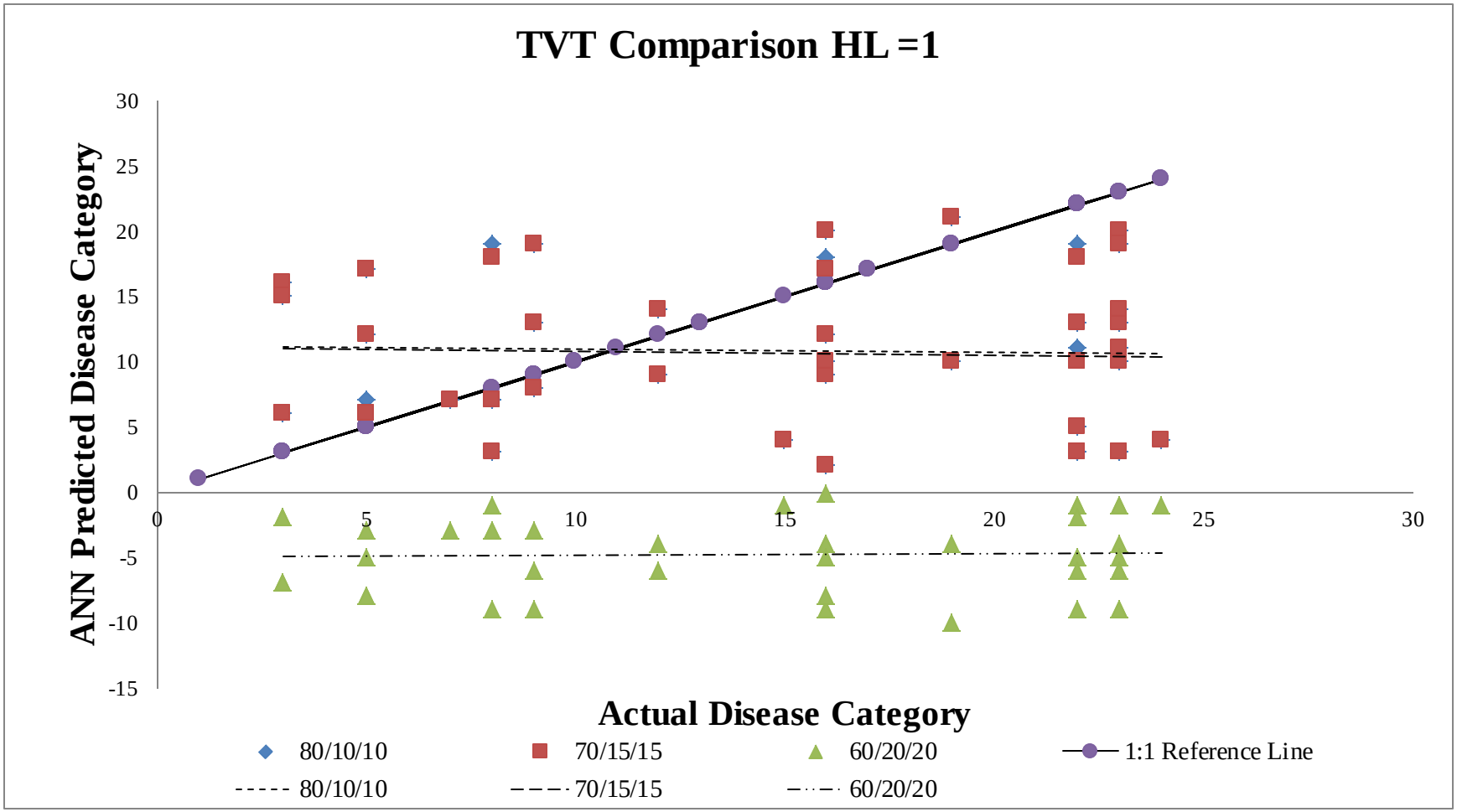


Figure 7: TVT Comparison HL =1

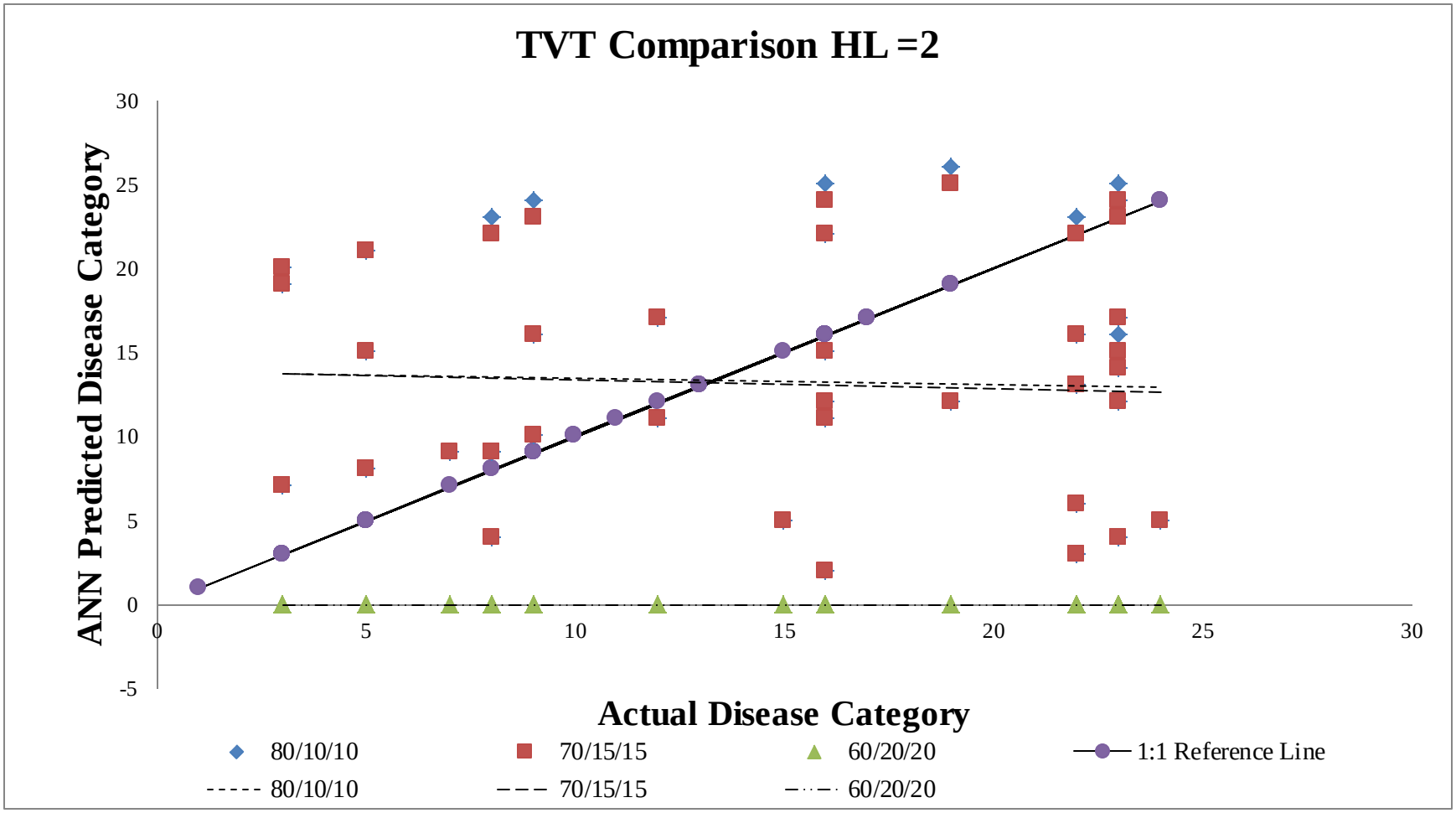


Figure 8: TVT Comparison HL =2

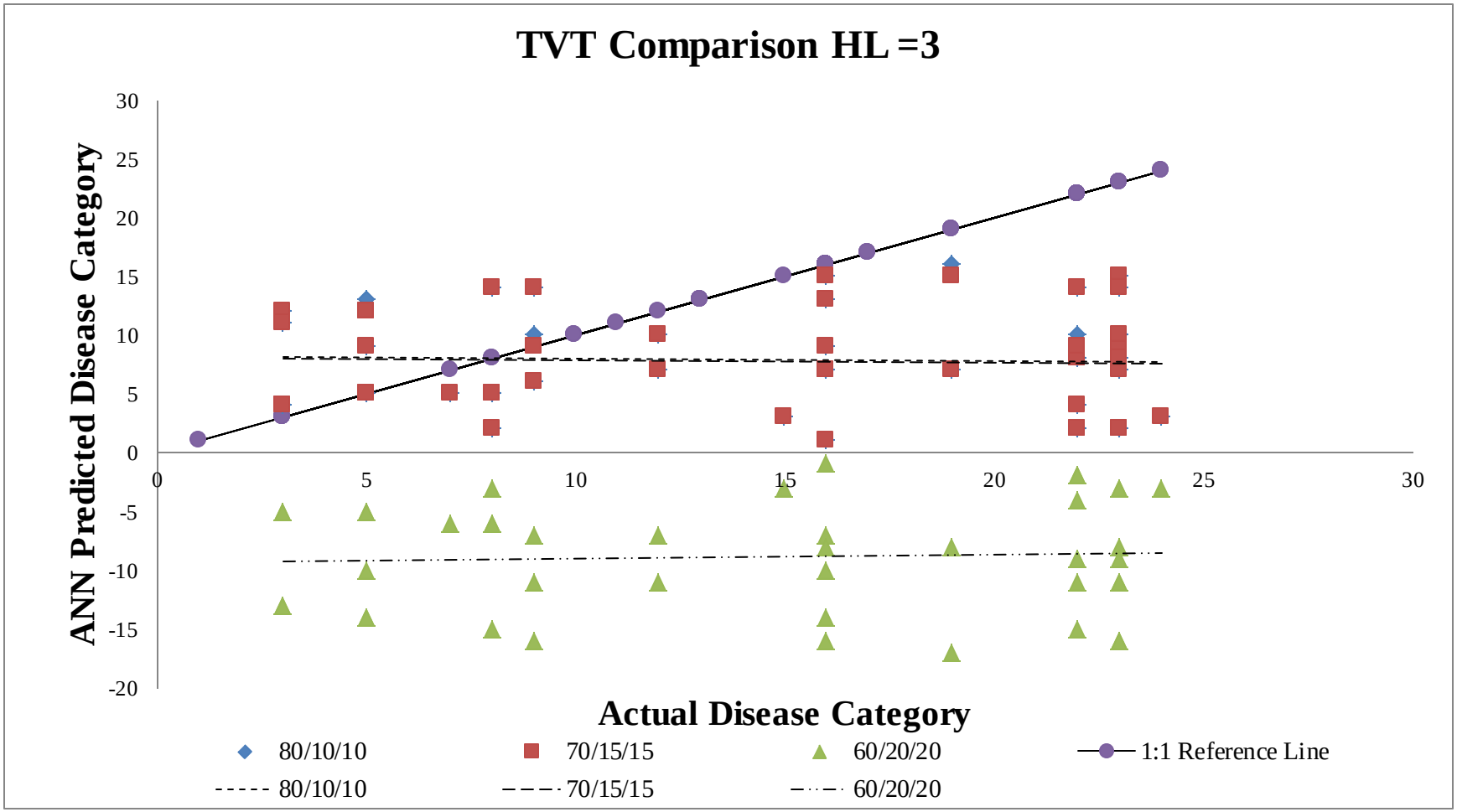


Figure 9: TVT Comparison HL =3

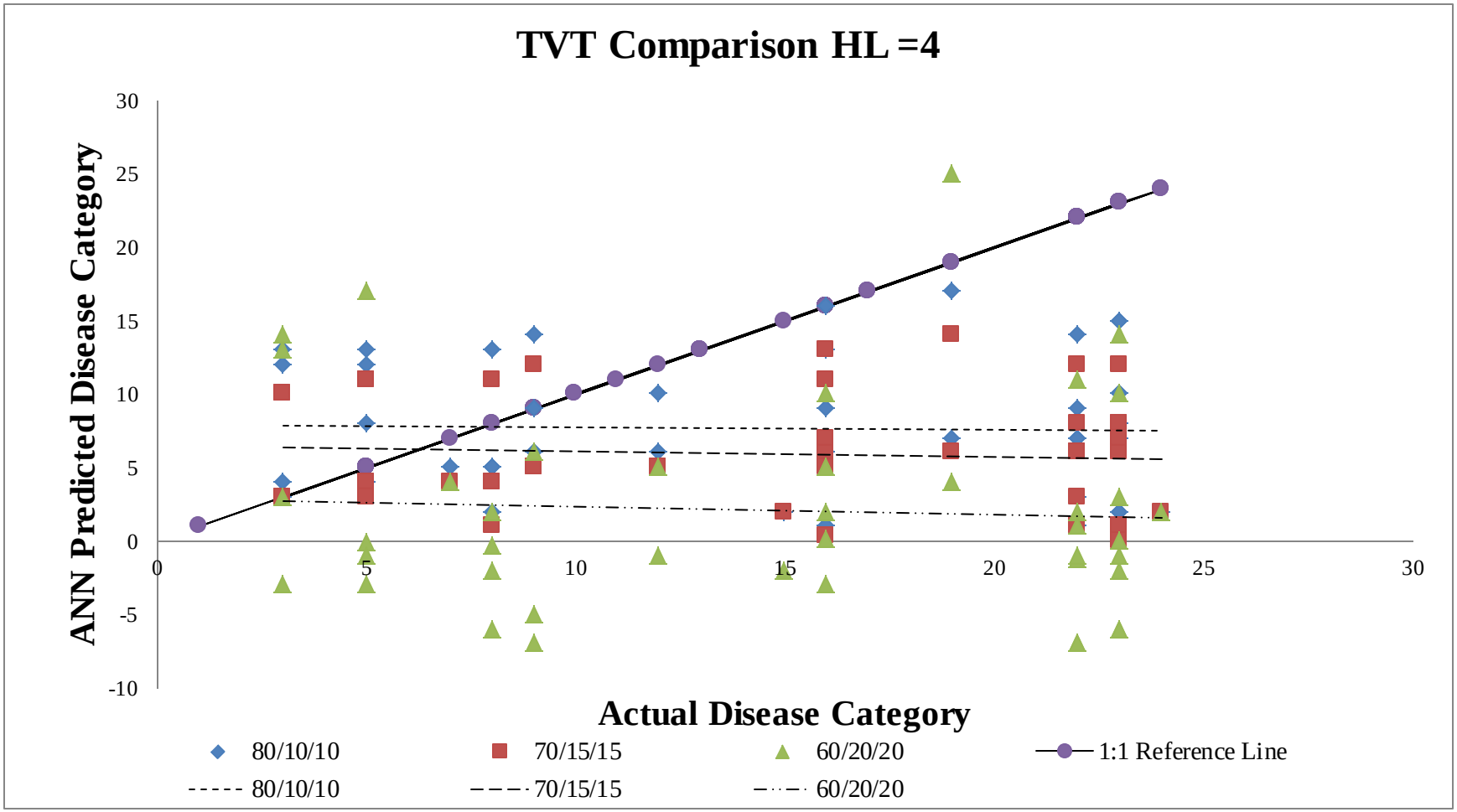


Figure 10: TVT Comparison HL =4

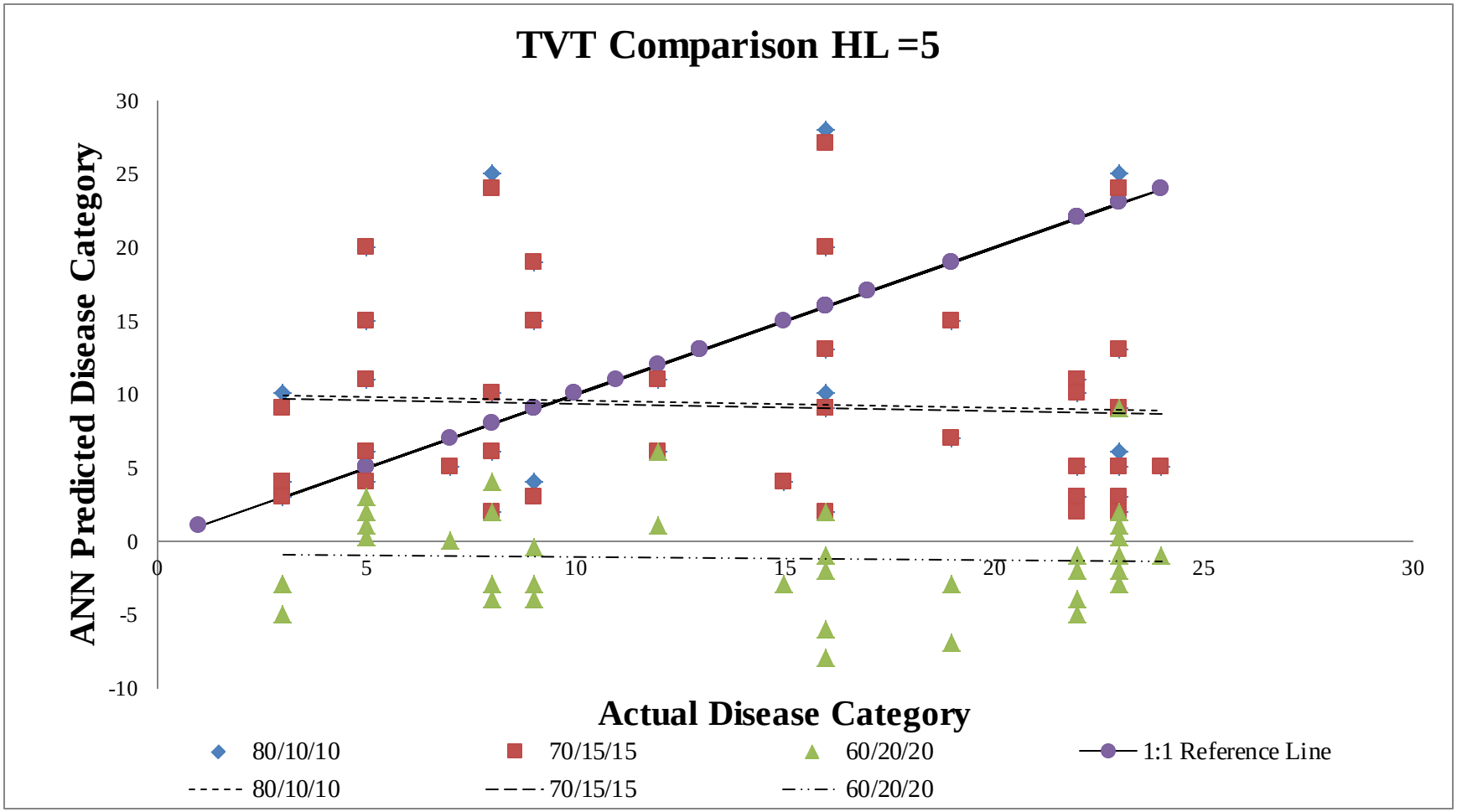


Figure 11: TVT Comparison HL =5

Appendix B

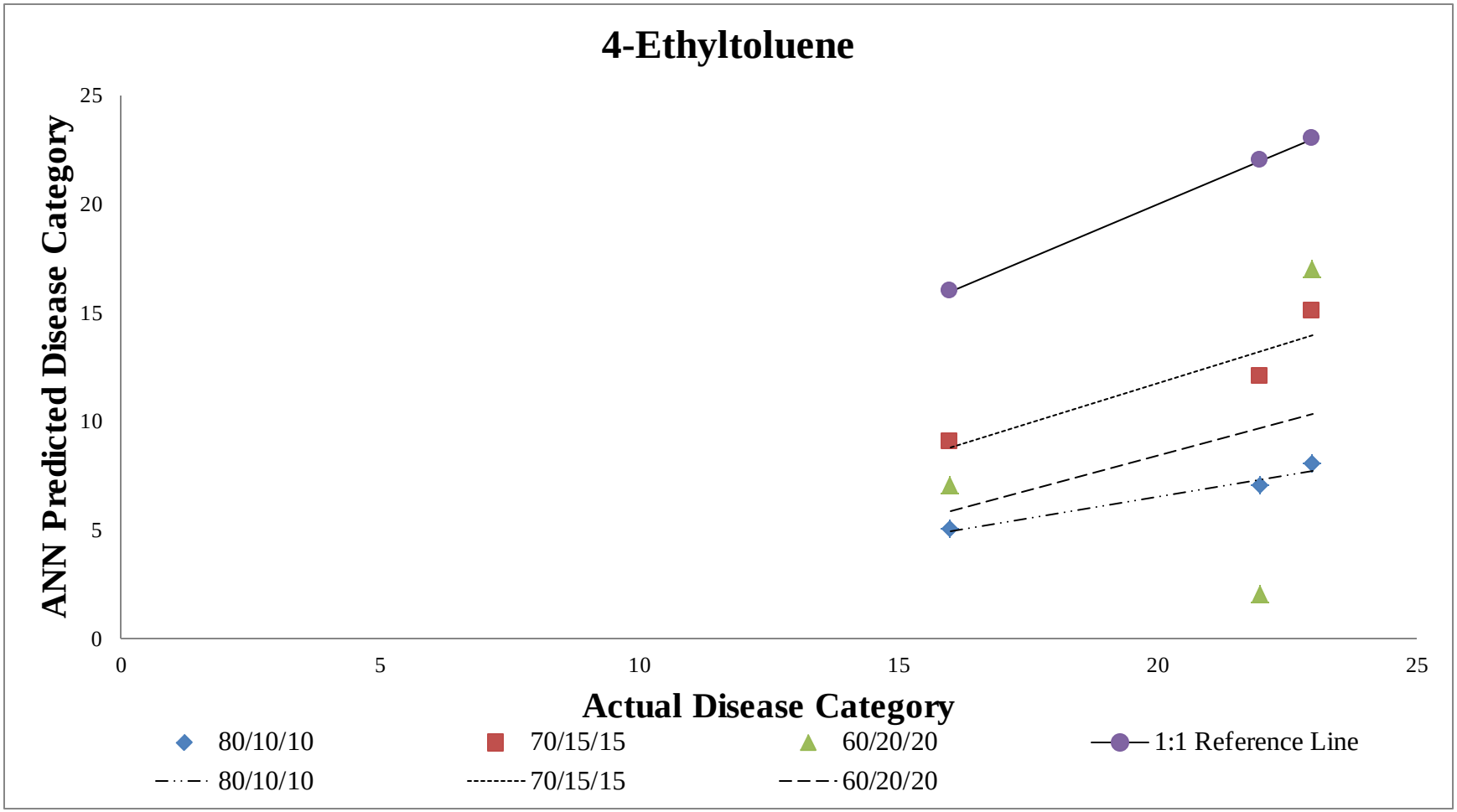


Figure 12: 4-Ethyltoluene

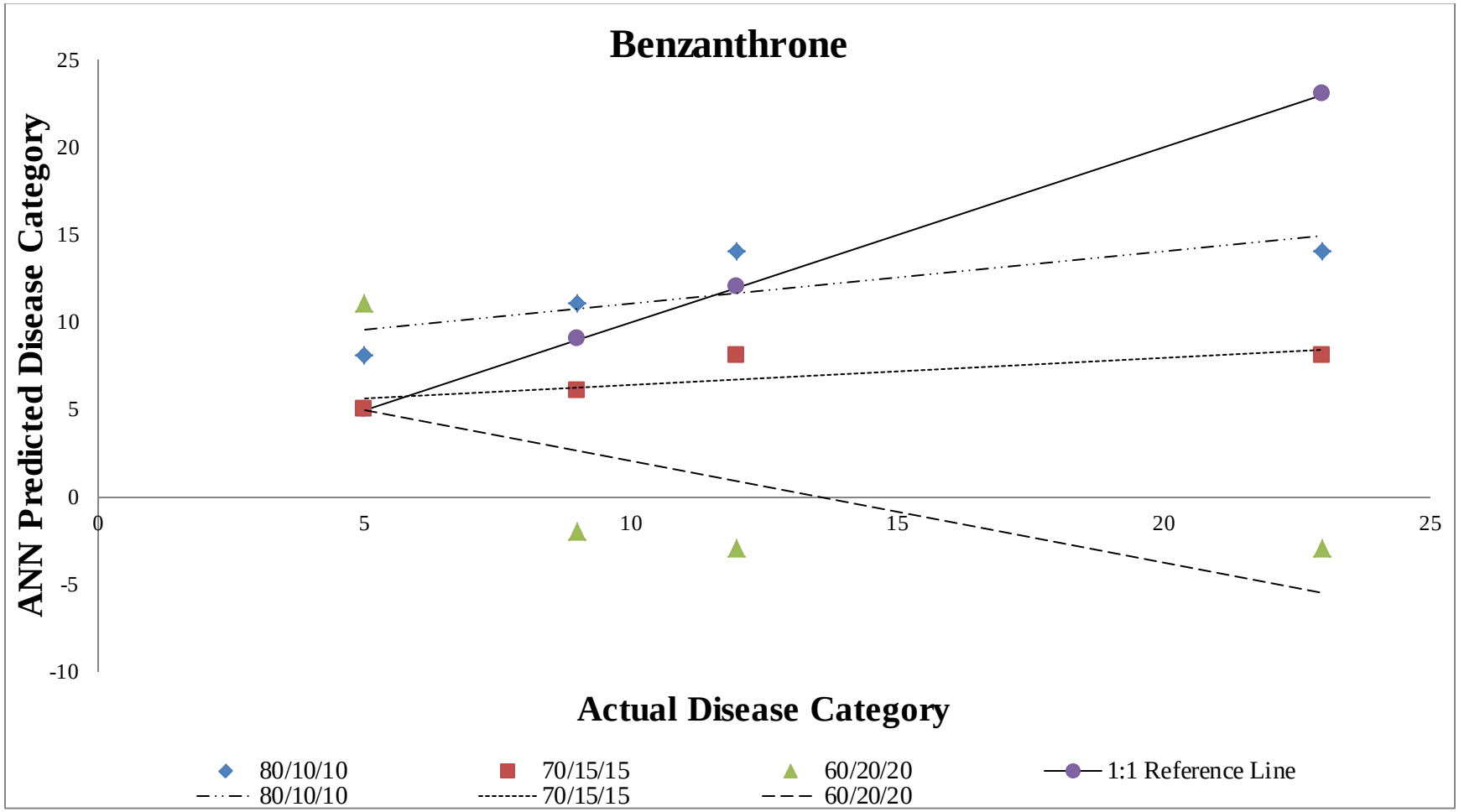


Figure 13: Benzanthrone

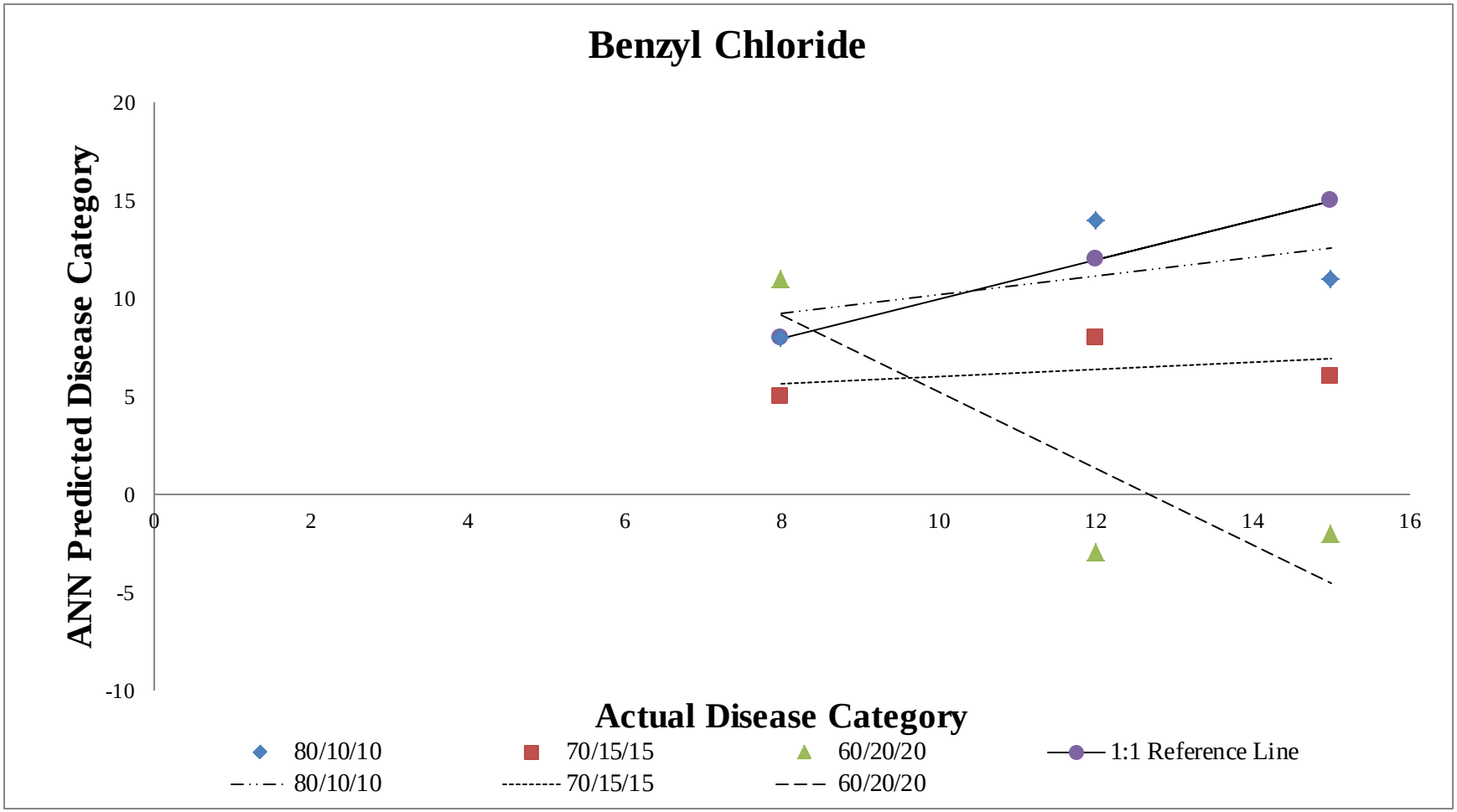


Figure 14: Benzyl Chloride

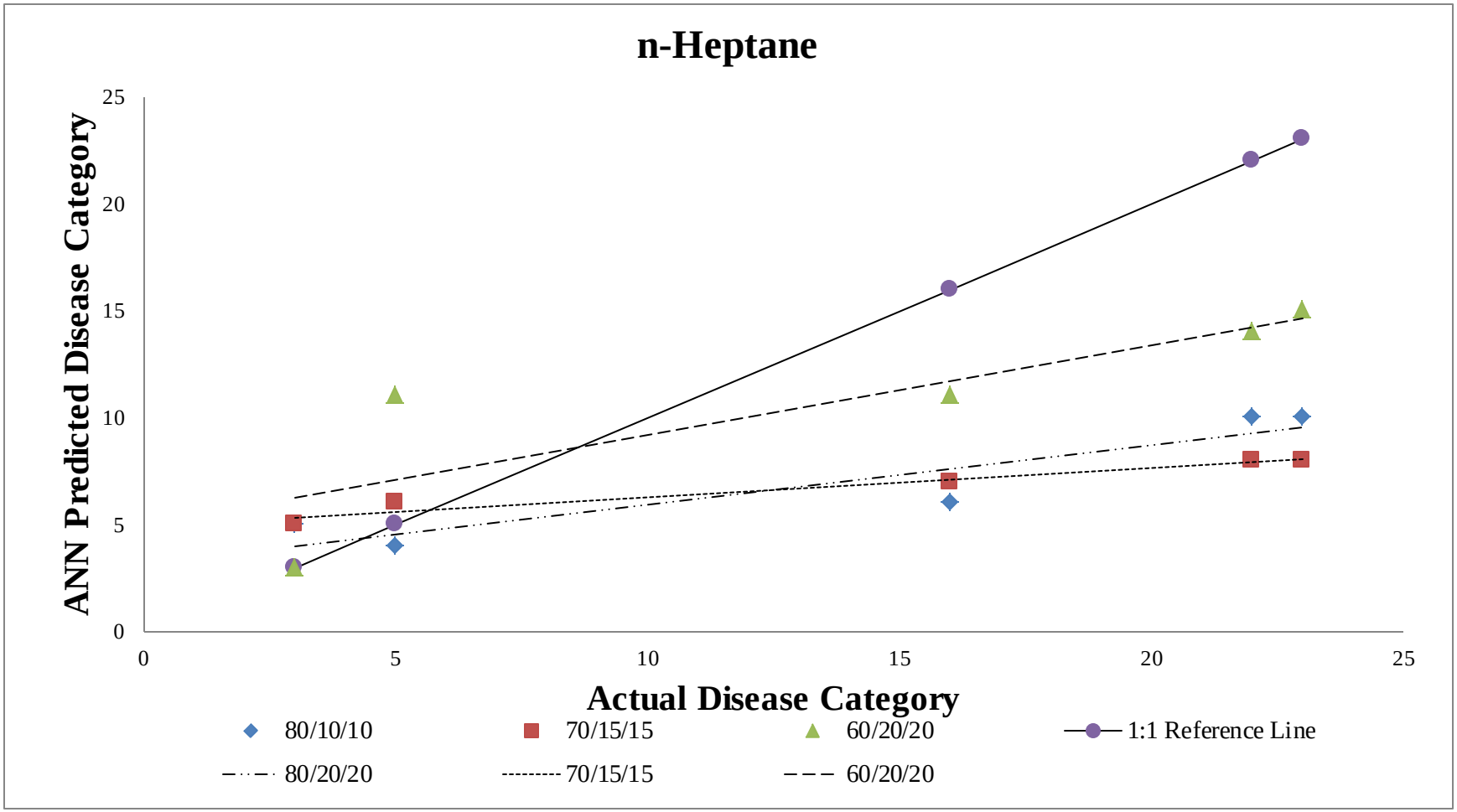


Figure 15: n-Heptane

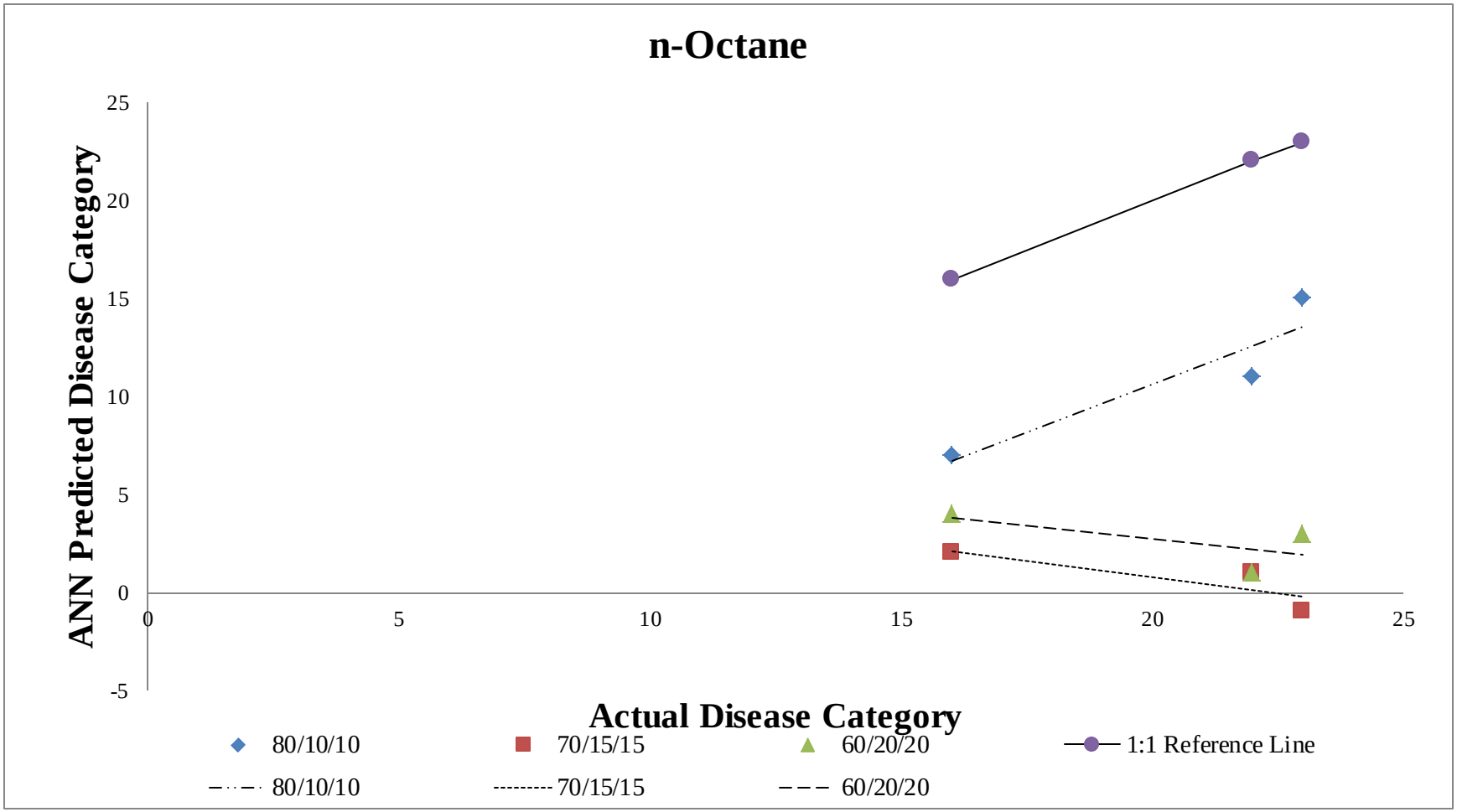


Figure 16: n-Octane

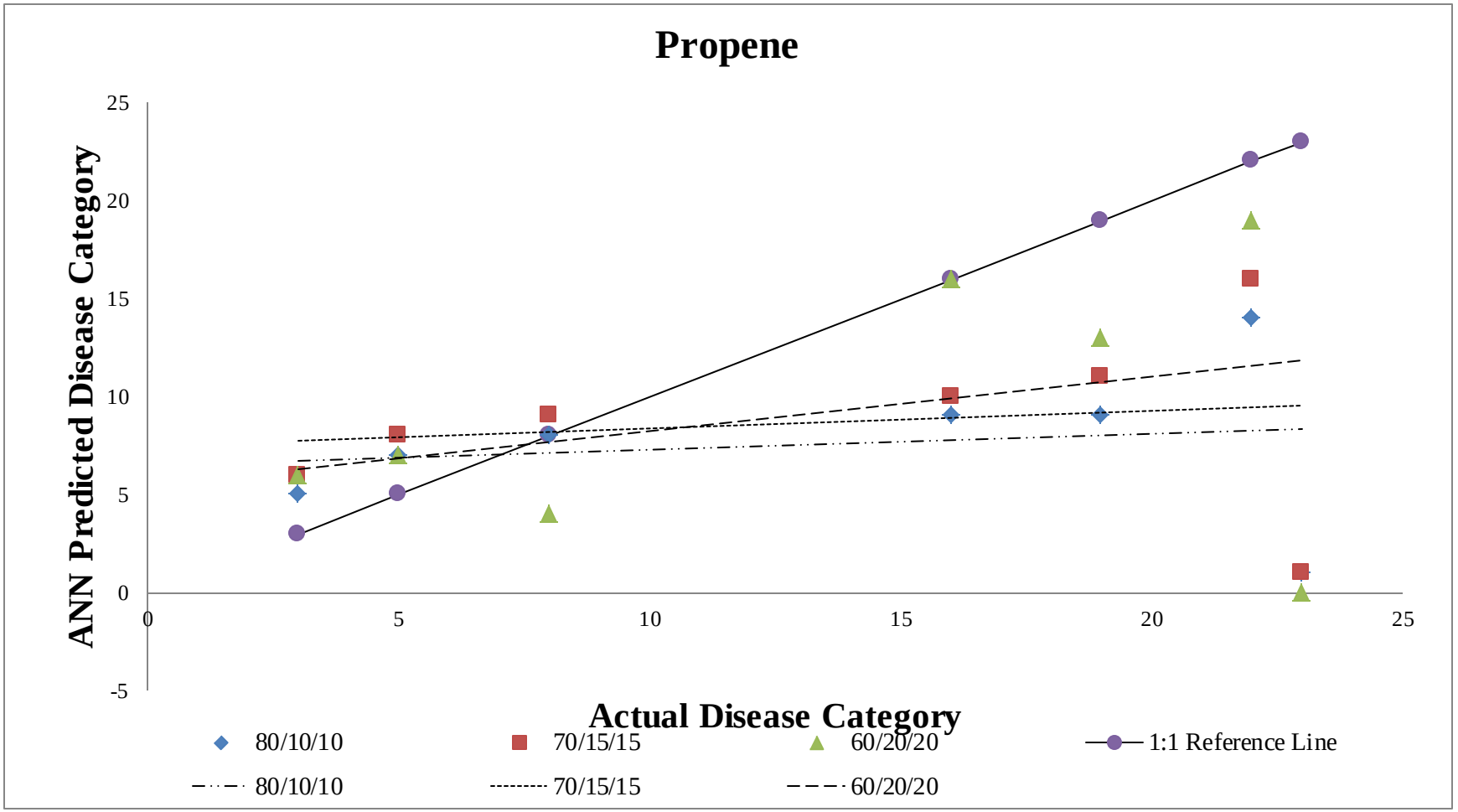


Figure 17: Propene

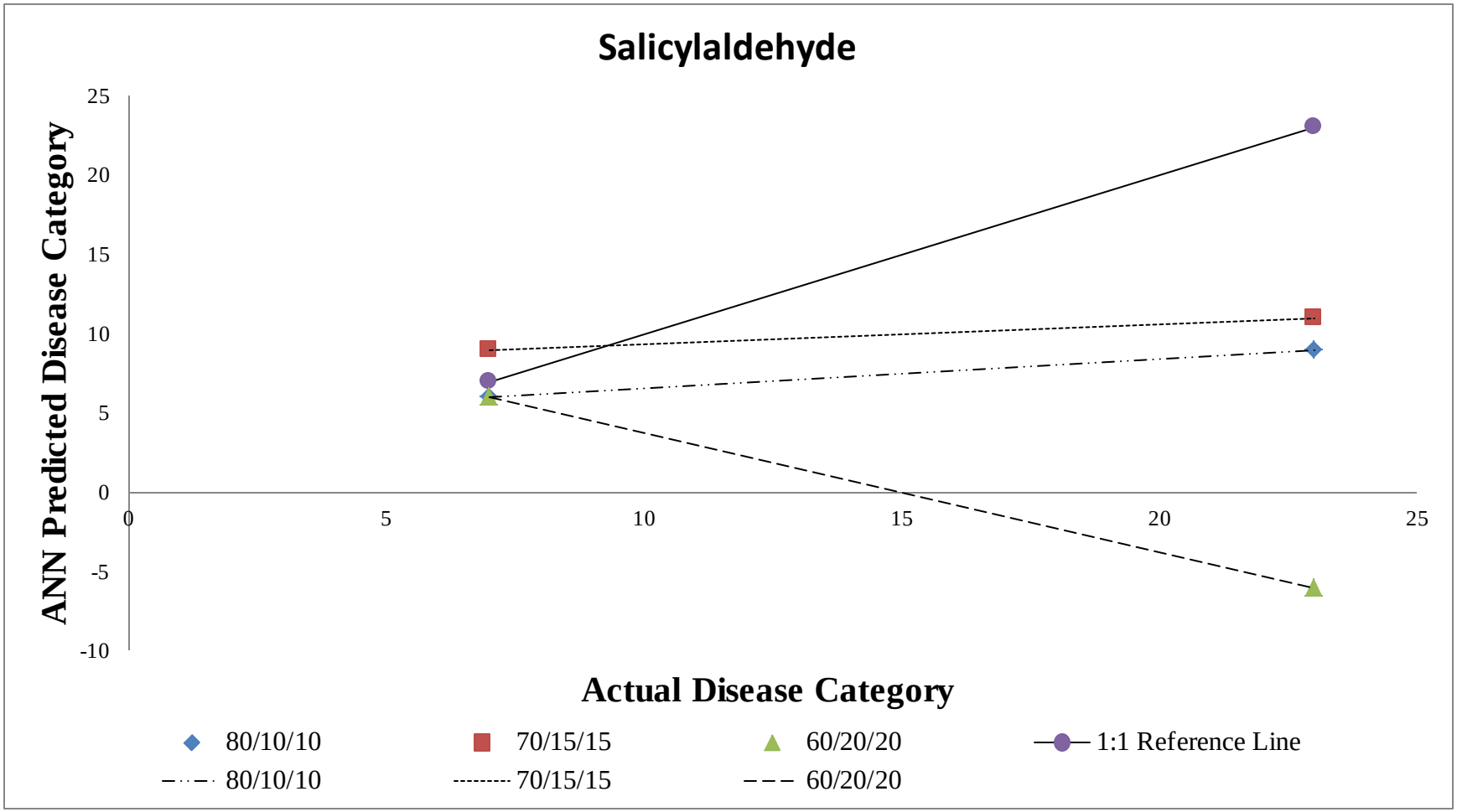


Figure 18: Salicylaldehyde

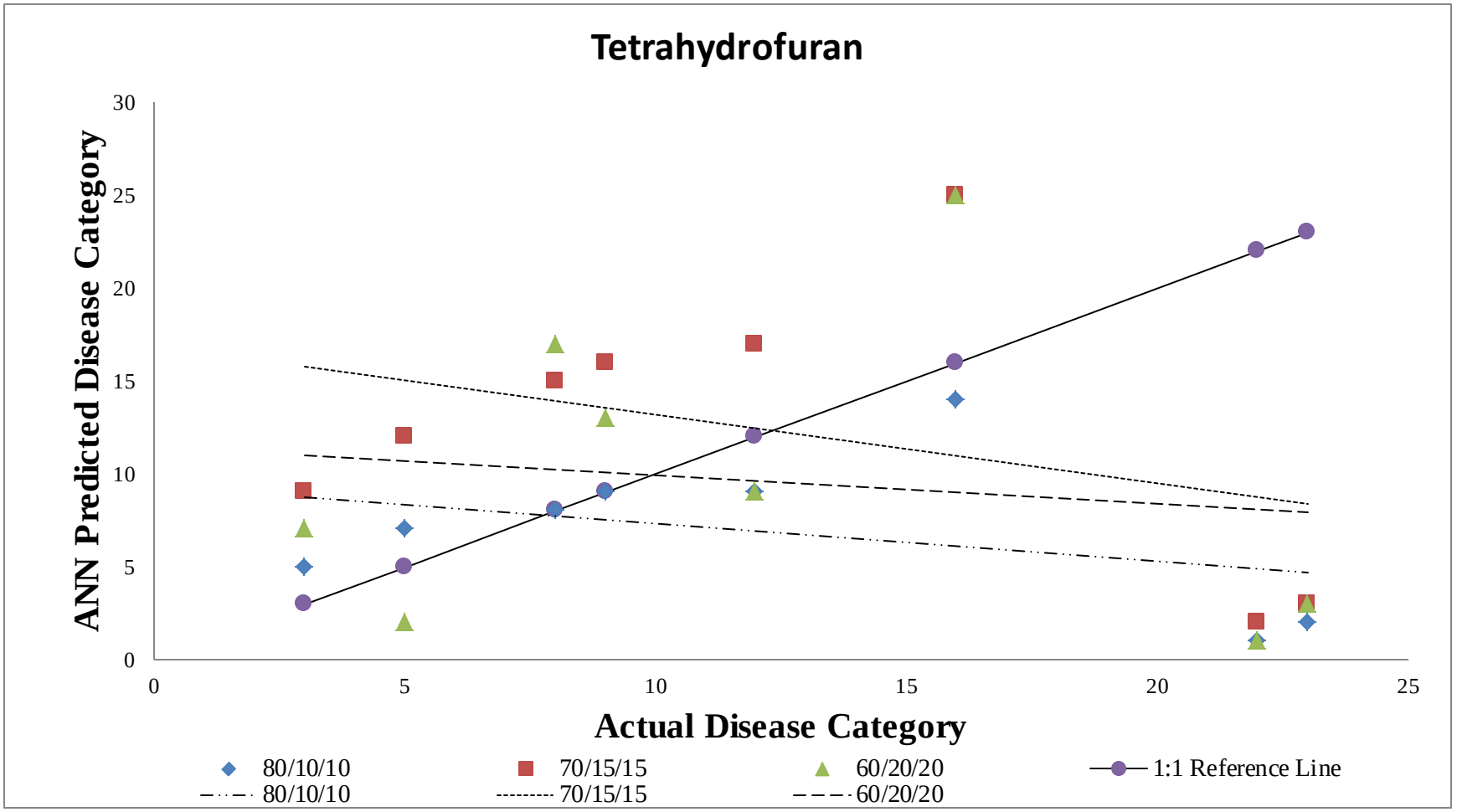


Figure 19: Tetrahydrofuran

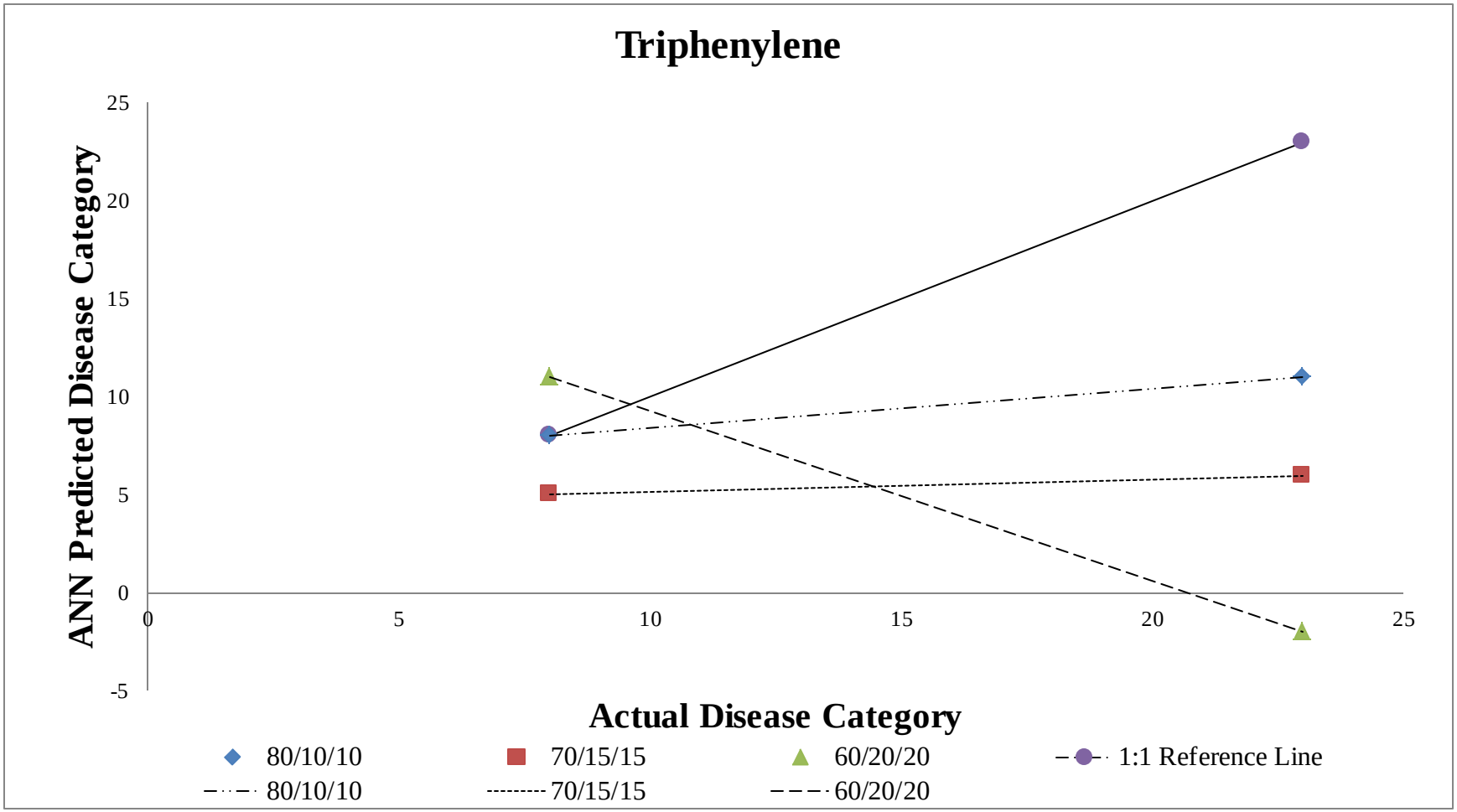


Figure 20: Triphenylene

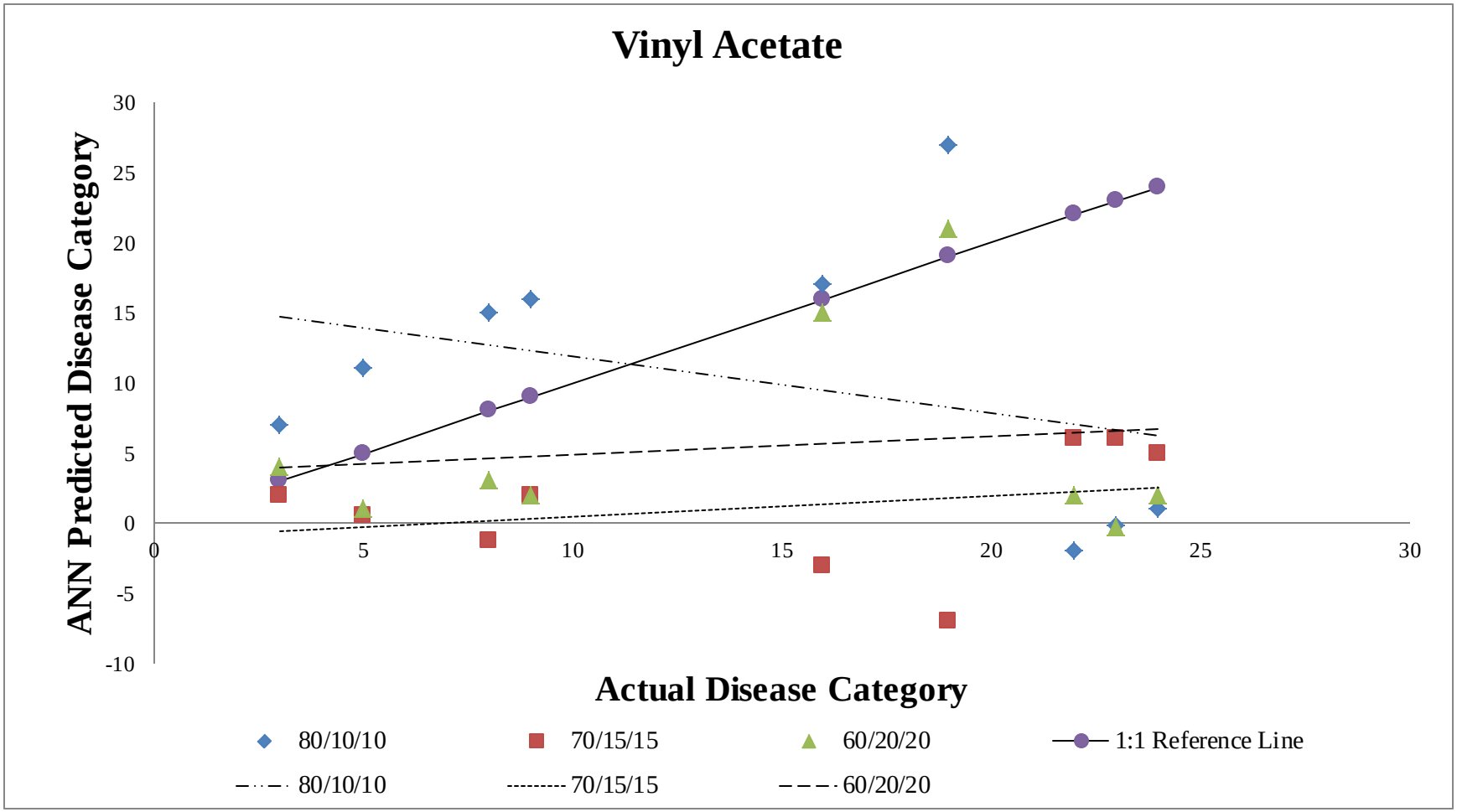


Figure 21: Vinyl Acetate

Appendix C

Table 15: ANN Training Data

Chemical Name	Molecular Weight (UU)	Hydrogen Acceptors (VV)	Hydrogen Donors (WW)	Disease Category (ZZ)
Acetone	58.08	1	0	9
	58.08	1	0	12
	58.08	1	0	15
	58.08	1	0	16
	58.08	1	0	17
	58.08	1	0	25
Aciclovir	225.21	8	4	2
	225.21	8	4	3
	225.21	8	4	4
	225.21	8	4	5
	225.21	8	4	8
	225.21	8	4	9
	225.21	8	4	11
	225.21	8	4	12
	225.21	8	4	13
	225.21	8	4	15
	225.21	8	4	16
	225.21	8	4	17
	225.21	8	4	19
	225.21	8	4	20
	225.21	8	4	21
	225.21	8	4	23
	225.21	8	4	24
	225.21	8	4	25
	225.21	8	4	26
Alprazolam	308.77	4	0	3
	308.77	4	0	4
	308.77	4	0	5
	308.77	4	0	7
	308.77	4	0	8
	308.77	4	0	9
	308.77	4	0	11
	308.77	4	0	12
	308.77	4	0	13
	308.77	4	0	16
	308.77	4	0	21
	308.77	4	0	23
	308.77	4	0	25
Ammonium Sulfate	132.14	4	2	3
	132.14	4	2	17
	132.14	4	2	22
Aspirin	180.16	4	1	2
	180.16	4	1	3
	180.16	4	1	4
	180.16	4	1	5
	180.16	4	1	6
	180.16	4	1	7
	180.16	4	1	8
	180.16	4	1	9
	180.16	4	1	10
	180.16	4	1	11
	180.16	4	1	12
	180.16	4	1	13
	180.16	4	1	14
	180.16	4	1	15
	180.16	4	1	16
	180.16	4	1	17

Aspirin (continued)	180.16	4	1	19
	180.16	4	1	21
	180.16	4	1	22
	180.16	4	1	23
	180.16	4	1	24
	180.16	4	1	25
	180.16	4	1	26
	180.16	4	1	27
Atenolol	266.34	5	4	3
	266.34	5	4	4
	266.34	5	4	5
	266.34	5	4	7
	266.34	5	4	9
	266.34	5	4	11
	266.34	5	4	12
	266.34	5	4	13
	266.34	5	4	14
	266.34	5	4	15
	266.34	5	4	16
	266.34	5	4	17
	266.34	5	4	21
	266.34	5	4	23
	266.34	5	4	24
	266.34	5	4	25
Azithromycin	749.00	14	5	2
	749.00	14	5	3
	749.00	14	5	5
	749.00	14	5	7
	749.00	14	5	8
	749.00	14	5	9
	749.00	14	5	10
	749.00	14	5	11
	749.00	14	5	12
	749.00	14	5	13
	749.00	14	5	14
	749.00	14	5	15
	749.00	14	5	16
	749.00	14	5	17
	749.00	14	5	19
	749.00	14	5	21
	749.00	14	5	22
	749.00	14	5	23
	749.00	14	5	24
	749.00	14	5	24
	749.00	14	5	26
Benzene	78.12	0	0	1
	78.12	0	0	3
	78.12	0	0	4
	78.12	0	0	5
	78.12	0	0	7
	78.12	0	0	9
	78.12	0	0	10
	78.12	0	0	11
	78.12	0	0	14
	78.12	0	0	15
	78.12	0	0	16
	78.12	0	0	17
	78.12	0	0	18
	78.12	0	0	21
	78.12	0	0	22
	78.12	0	0	23
	78.12	0	0	25
	78.12	0	0	27
Benzyl-penicillin	334.40	6	2	1

Benzyl-penicillin (continued)	334.40	6	2	2
	334.40	6	2	3
	334.40	6	2	4
	334.40	6	2	5
	334.40	6	2	7
	334.40	6	2	8
	334.40	6	2	9
	334.40	6	2	10
	334.40	6	2	11
	334.40	6	2	12
	334.40	6	2	14
	334.40	6	2	16
	334.40	6	2	17
	334.40	6	2	19
	334.40	6	2	20
	334.40	6	2	21
	334.40	6	2	22
	334.40	6	2	23
	334.40	6	2	24
	334.40	6	2	25
	334.40	6	2	26
Caffeine	194.19	6	0	1
	194.19	6	0	2
	194.19	6	0	3
	194.19	6	0	4
	194.19	6	0	5
	194.19	6	0	7
	194.19	6	0	8
	194.19	6	0	9
	194.19	6	0	10
	194.19	6	0	11
	194.19	6	0	12
	194.19	6	0	13
	194.19	6	0	14
	194.19	6	0	15
	194.19	6	0	16
	194.19	6	0	17
	194.19	6	0	19
	194.19	6	0	21
	194.19	6	0	22
	194.19	6	0	23
	194.19	6	0	25
	194.19	6	0	27
Candoxatril	515.65	8	2	3
	515.65	8	2	7
	515.65	8	2	16
	515.65	8	2	17
Carbamazepine	236.28	3	2	1
	236.28	3	2	3
	236.28	3	2	4
	236.28	3	2	5
	236.28	3	2	7
	236.28	3	2	8
	236.28	3	2	9
	236.28	3	2	10
	236.28	3	2	11
	236.28	3	2	12
	236.28	3	2	13
	236.28	3	2	14
	236.28	3	2	15
	236.28	3	2	16
	236.28	3	2	17
	236.28	3	2	19
	236.28	3	2	21

Carbamazepine (continued)	236.28	3	2	22
	236.28	3	2	23
	236.28	3	2	24
	236.28	3	2	25
	236.28	3	2	26
Caustic Soda	40.00	1	1	1
	40.00	1	1	5
	40.00	1	1	19
	40.00	1	1	22
	40.00	1	1	27
Chloramphenicol	323.14	7	3	2
	323.14	7	3	3
	323.14	7	3	5
	323.14	7	3	8
	323.14	7	3	9
	323.14	7	3	10
	323.14	7	3	11
	323.14	7	3	12
	323.14	7	3	14
	323.14	7	3	15
	323.14	7	3	16
	323.14	7	3	17
	323.14	7	3	19
	323.14	7	3	21
	323.14	7	3	22
	323.14	7	3	23
	323.14	7	3	25
Cimetidine	252.34	6	3	2
	252.34	6	3	3
	252.34	6	3	4
	252.34	6	3	5
	252.34	6	3	7
	252.34	6	3	8
	252.34	6	3	9
	252.34	6	3	11
	252.34	6	3	12
	252.34	6	3	13
	252.34	6	3	14
	252.34	6	3	15
	252.34	6	3	16
	252.34	6	3	17
	252.34	6	3	19
	252.34	6	3	21
	252.34	6	3	22
	252.34	6	3	23
	252.34	6	3	24
	252.34	6	3	25
	252.34	6	3	26
	252.34	6	3	27
Clonidine	230.10	3	2	3
	230.10	3	2	4
	230.10	3	2	5
	230.10	3	2	7
	230.10	3	2	8
	230.10	3	2	9
	230.10	3	2	11
	230.10	3	2	12
	230.10	3	2	13
	230.10	3	2	14
	230.10	3	2	16
	230.10	3	2	17
	230.10	3	2	19
	230.10	3	2	21
	230.10	3	2	22

Clonidine (continued)	230.10	3	2	23
	230.10	3	2	24
	230.10	3	2	25
Copper Sulfate	159.61	4	0	4
	159.61	4	0	5
	159.61	4	0	9
	159.61	4	0	12
	159.61	4	0	13
	159.61	4	0	14
	159.61	4	0	16
	159.61	4	0	17
	159.61	4	0	19
	159.61	4	0	21
	159.61	4	0	22
	159.61	4	0	25
Cyclosporine	1202.64	23	5	1
	1202.64	23	5	2
	1202.64	23	5	3
	1202.64	23	5	4
	1202.64	23	5	5
	1202.64	23	5	7
	1202.64	23	5	8
	1202.64	23	5	9
	1202.64	23	5	10
	1202.64	23	5	11
	1202.64	23	5	12
	1202.64	23	5	13
	1202.64	23	5	14
	1202.64	23	5	15
	1202.64	23	5	16
	1202.64	23	5	17
	1202.64	23	5	19
	1202.64	23	5	20
	1202.64	23	5	21
	1202.64	23	5	22
	1202.64	23	5	23
	1202.64	23	5	25
	1202.64	23	5	26
Desipramine	266.39	2	1	2
	266.39	2	1	3
	266.39	2	1	4
	266.39	2	1	5
	266.39	2	1	7
	266.39	2	1	8
	266.39	2	1	9
	266.39	2	1	11
	266.39	2	1	12
	266.39	2	1	13
	266.39	2	1	15
	266.39	2	1	16
	266.39	2	1	17
	266.39	2	1	21
	266.39	2	1	22
	266.39	2	1	23
	266.39	2	1	25
Dexamethasone	392.47	5	3	2
	392.47	5	3	3
	392.47	5	3	4
	392.47	5	3	5
	392.47	5	3	7
	392.47	5	3	8
	392.47	5	3	9
	392.47	5	3	10
	392.47	5	3	11

Dexamethasone (continued)	392.47	5	3	12
	392.47	5	3	13
	392.47	5	3	14
	392.47	5	3	15
	392.47	5	3	16
	392.47	5	3	17
	392.47	5	3	19
	392.47	5	3	20
	392.47	5	3	21
	392.47	5	3	22
	392.47	5	3	23
	392.47	5	3	24
	392.47	5	3	25
	392.47	5	3	26
	392.47	5	3	27
Diazepam	284.75	3	0	2
	284.75	3	0	2
	284.75	3	0	4
	284.75	3	0	5
	284.75	3	0	8
	284.75	3	0	9
	284.75	3	0	11
	284.75	3	0	12
	284.75	3	0	13
	284.75	3	0	14
	284.75	3	0	15
	284.75	3	0	16
	284.75	3	0	17
	284.75	3	0	19
	284.75	3	0	21
	284.75	3	0	22
	284.75	3	0	23
	284.75	3	0	24
	284.75	3	0	25
	284.75	3	0	27
Diclofenac	296.15	3	2	2
	296.15	3	2	3
	296.15	3	2	4
	296.15	3	2	5
	296.15	3	2	7
	296.15	3	2	8
	296.15	3	2	9
	296.15	3	2	11
	296.15	3	2	12
	296.15	3	2	13
	296.15	3	2	14
	296.15	3	2	15
	296.15	3	2	16
	296.15	3	2	17
	296.15	3	2	19
	296.15	3	2	21
	296.15	3	2	22
	296.15	3	2	23
	296.15	3	2	25
	296.15	3	2	26
	296.15	3	2	27
Diltiazem-HCl	414.53	6	0	3
	414.53	6	0	4
	414.53	6	0	4
	414.53	6	0	7
	414.53	6	0	8
	414.53	6	0	9
	414.53	6	0	10
	414.53	6	0	11

Diltiazem-HCl (continued)	414.53	6	0	12
	414.53	6	0	13
	414.53	6	0	14
	414.53	6	0	15
	414.53	6	0	16
	414.53	6	0	17
	414.53	6	0	21
	414.53	6	0	23
	414.53	6	0	25
Doxorubicin	543.53	12	7	1
	543.53	12	7	2
	543.53	12	7	3
	543.53	12	7	4
	543.53	12	7	5
	543.53	12	7	7
	543.53	12	7	8
	543.53	12	7	9
	543.53	12	7	10
	543.53	12	7	11
	543.53	12	7	12
	543.53	12	7	13
	543.53	12	7	14
	543.53	12	7	15
	543.53	12	7	16
	543.53	12	7	17
	543.53	12	7	19
	543.53	12	7	21
	543.53	12	7	22
	543.53	12	7	23
	543.53	12	7	24
	543.53	12	7	25
	543.53	12	7	26
	543.53	12	7	27
Enalaprilat	376.46	7	2	3
	376.46	7	2	5
	376.46	7	2	21
Erythromycin	733.95	14	5	1
	733.95	14	5	2
	733.95	14	5	3
	733.95	14	5	4
	733.95	14	5	5
	733.95	14	5	7
	733.95	14	5	8
	733.95	14	5	9
	733.95	14	5	10
	733.95	14	5	11
	733.95	14	5	12
	733.95	14	5	13
	733.95	14	5	14
	733.95	14	5	15
	733.95	14	5	16
	733.95	14	5	17
	733.95	14	5	19
	733.95	14	5	20
	733.95	14	5	21
	733.95	14	5	22
	733.95	14	5	23
	733.95	14	5	24
	733.95	14	5	25
	733.95	14	5	26
Ethylene Glycol	62.07	2	2	4
	62.07	2	2	5
	62.07	2	2	9
	62.07	2	2	12

Ethylene Glycol (continued)	62.07	2	2	14
	62.07	2	2	16
	62.07	2	2	25
Famotidine	337.45	9	8	3
	337.45	9	8	4
	337.45	9	8	5
	337.45	9	8	7
	337.45	9	8	9
	337.45	9	8	11
	337.45	9	8	12
	337.45	9	8	13
	337.45	9	8	14
	337.45	9	8	15
	337.45	9	8	16
	337.45	9	8	17
	337.45	9	8	21
	337.45	9	8	22
	337.45	9	8	23
	337.45	9	8	25
	337.45	9	8	26
	337.45	9	8	27
Felodipine	384.26	5	1	3
	384.26	5	1	9
	384.26	5	1	12
	384.26	5	1	16
	384.26	5	1	21
	384.26	5	1	23
	384.26	5	1	24
Ferric Chloride	162.20	0	0	3
	162.20	0	0	5
	162.20	0	0	9
	162.20	0	0	11
	162.20	0	0	12
	162.20	0	0	16
	162.20	0	0	21
	162.20	0	0	25
	162.20	0	0	27
Fluorouracil	130.08	4	2	2
	130.08	4	2	3
	130.08	4	2	4
	130.08	4	2	5
	130.08	4	2	7
	130.08	4	2	8
	130.08	4	2	9
	130.08	4	2	10
	130.08	4	2	11
	130.08	4	2	12
	130.08	4	2	13
	130.08	4	2	14
	130.08	4	2	15
	130.08	4	2	16
	130.08	4	2	17
	130.08	4	2	19
	130.08	4	2	21
	130.08	4	2	22
	130.08	4	2	23
	130.08	4	2	24
	130.08	4	2	25
	130.08	4	2	26
	130.08	4	2	27
Flurbiprofen	244.27	2	1	3
	244.27	2	1	4
	244.27	2	1	5
	244.27	2	1	8

Flurbiprofen (continued)	244.27	2	1	9
	244.27	2	1	11
	244.27	2	1	12
	244.27	2	1	13
	244.27	2	1	14
	244.27	2	1	15
	244.27	2	1	16
	244.27	2	1	17
	244.27	2	1	19
	244.27	2	1	21
	244.27	2	1	22
	244.27	2	1	23
	244.27	2	1	24
	244.27	2	1	25
Formaldehyde	30.03	1	0	1
	30.03	1	0	1
	30.03	1	0	2
	30.03	1	0	3
	30.03	1	0	4
	30.03	1	0	5
	30.03	1	0	7
	30.03	1	0	8
	30.03	1	0	9
	30.03	1	0	11
	30.03	1	0	12
	30.03	1	0	13
	30.03	1	0	14
	30.03	1	0	15
	30.03	1	0	16
	30.03	1	0	18
	30.03	1	0	19
	30.03	1	0	20
	30.03	1	0	21
	30.03	1	0	22
	30.03	1	0	23
	30.03	1	0	24
	30.03	1	0	25
	30.03	1	0	26
Furosemide	330.75	7	4	3
	330.75	7	4	4
	330.75	7	4	5
	330.75	7	4	7
	330.75	7	4	9
	330.75	7	4	10
	330.75	7	4	11
	330.75	7	4	12
	330.75	7	4	13
	330.75	7	4	14
	330.75	7	4	15
	330.75	7	4	16
	330.75	7	4	17
	330.75	7	4	19
	330.75	7	4	21
	330.75	7	4	22
	330.75	7	4	23
	330.75	7	4	25
Gabapentin	30.03	1	0	27
	171.24	3	2	3
	171.24	3	2	4
	171.24	3	2	5
	171.24	3	2	7
	171.24	3	2	8
	171.24	3	2	9
	171.24	3	2	12

Gabapentin (continued)	171.24	3	2	13
	171.24	3	2	15
	171.24	3	2	16
	171.24	3	2	17
	171.24	3	2	21
	171.24	3	2	23
	171.24	3	2	24
Glycerol	92.09	3	3	1
	92.09	3	3	3
	92.09	3	3	7
	92.09	3	3	8
	92.09	3	3	9
	92.09	3	3	12
	92.09	3	3	14
	92.09	3	3	16
	92.09	3	3	17
	92.09	3	3	21
Hydrochloric Acid	36.46	0	1	2
	36.46	0	1	3
	36.46	0	1	4
	36.46	0	1	5
	36.46	0	1	7
	36.46	0	1	9
	36.46	0	1	11
	36.46	0	1	12
	36.46	0	1	13
	36.46	0	1	14
	36.46	0	1	15
	36.46	0	1	16
	36.46	0	1	17
	36.46	0	1	18
	36.46	0	1	19
	36.46	0	1	20
	36.46	0	1	21
	36.46	0	1	22
	36.46	0	1	23
	36.46	0	1	24
	36.46	0	1	25
	36.46	0	1	27
Hydrochlorothiazide	297.74	7	4	3
	297.74	7	4	4
	297.74	7	4	5
	297.74	7	4	7
	297.74	7	4	9
	297.74	7	4	11
	297.74	7	4	12
	297.74	7	4	13
	297.74	7	4	14
	297.74	7	4	16
	297.74	7	4	17
	297.74	7	4	21
	297.74	7	4	22
	297.74	7	4	23
	297.74	7	4	24
	297.74	7	4	25
Hydrofluoric Acid	20.01	1	1	5
	20.01	1	1	9
	20.01	1	1	12
	20.01	1	1	13
	20.01	1	1	14
	20.01	1	1	16
	20.01	1	1	17
	20.01	1	1	21
	20.01	1	1	22

Hydrofluoric Acid (continued)	20.01	1	1	24
	20.01	1	1	25
	20.01	1	1	27
Ibuprofen	206.29	2	1	2
	206.29	2	1	3
	206.29	2	1	4
	206.29	2	1	5
	206.29	2	1	7
	206.29	2	1	8
	206.29	2	1	9
	206.29	2	1	10
	206.29	2	1	11
	206.29	2	1	12
	206.29	2	1	13
	206.29	2	1	14
	206.29	2	1	15
	206.29	2	1	16
	206.29	2	1	17
	206.29	2	1	19
	206.29	2	1	21
	206.29	2	1	22
	206.29	2	1	23
	206.29	2	1	24
	206.29	2	1	25
	206.29	2	1	26
Imipramine	280.42	2	0	3
	280.42	2	0	4
	280.42	2	0	5
	280.42	2	0	7
	280.42	2	0	8
	280.42	2	0	9
	280.42	2	0	11
	280.42	2	0	12
	280.42	2	0	13
	280.42	2	0	14
	280.42	2	0	15
	280.42	2	0	16
	280.42	2	0	17
	280.42	2	0	19
	280.42	2	0	21
	280.42	2	0	22
	280.42	2	0	23
	280.42	2	0	24
	280.42	2	0	25
Isopropyl Alcohol	60.10	1	1	2
Itraconazole	705.65	12	0	2
	705.65	12	0	3
	705.65	12	0	4
	705.65	12	0	5
	705.65	12	0	7
	705.65	12	0	9
	705.65	12	0	11
	705.65	12	0	12
	705.65	12	0	13
	705.65	12	0	14
	705.65	12	0	15
	705.65	12	0	16
	705.65	12	0	17
	705.65	12	0	19
	705.65	12	0	20
	705.65	12	0	21
	705.65	12	0	22
	705.65	12	0	23
	705.65	12	0	25

Ketoconazole	380.92	1	0	1
	380.92	1	0	2
	380.92	1	0	3
	380.92	1	0	4
	380.92	1	0	5
	380.92	1	0	7
	380.92	1	0	8
	380.92	1	0	9
	380.92	1	0	11
	380.92	1	0	12
	380.92	1	0	13
	380.92	1	0	14
	380.92	1	0	15
	380.92	1	0	16
	380.92	1	0	17
	380.92	1	0	20
	380.92	1	0	21
	380.92	1	0	22
	380.92	1	0	23
	380.92	1	0	24
	380.92	1	0	25
	380.92	1	0	27
Ketoprofen	254.29	3	1	3
	254.29	3	1	4
	254.29	3	1	5
	254.29	3	1	9
	254.29	3	1	11
	254.29	3	1	12
	254.29	3	1	13
	254.29	3	1	14
	254.29	3	1	16
	254.29	3	1	21
	254.29	3	1	22
	254.29	3	1	23
	254.29	3	1	25
	254.29	3	1	27
Labetalol-HCl	328.42	5	5	2
	328.42	5	5	3
	328.42	5	5	4
	328.42	5	5	4
	328.42	5	5	5
	328.42	5	5	7
	328.42	5	5	9
	328.42	5	5	11
	328.42	5	5	12
	328.42	5	5	13
	328.42	5	5	14
	328.42	5	5	15
	328.42	5	5	16
	328.42	5	5	17
	328.42	5	5	21
	328.42	5	5	22
	328.42	5	5	23
	328.42	5	5	25
Lisinopril	405.50	8	5	3
	405.50	8	5	4
	405.50	8	5	5
	405.50	8	5	7
	405.50	8	5	8
	405.50	8	5	9
	405.50	8	5	11
	405.50	8	5	12
	405.50	8	5	13
	405.50	8	5	14

Lisinopril (continued)	405.50	8	5	16
	405.50	8	5	17
	405.50	8	5	21
	405.50	8	5	22
	405.50	8	5	23
	405.50	8	5	24
	405.50	8	5	25
Magnesium Sulfate	120.37	4	0	2
	120.37	4	0	3
	120.37	4	0	4
	120.37	4	0	5
	120.37	4	0	8
	120.37	4	0	9
	120.37	4	0	11
	120.37	4	0	12
	120.37	4	0	13
	120.37	4	0	14
	120.37	4	0	16
	120.37	4	0	17
	120.37	4	0	21
	120.37	4	0	22
	120.37	4	0	25
	120.37	4	0	27
Mannitol	182.18	6	6	3
	182.18	6	6	5
	182.18	6	6	7
	182.18	6	6	8
	182.18	6	6	9
	182.18	6	6	11
	182.18	6	6	12
	182.18	6	6	14
	182.18	6	6	15
	182.18	6	6	16
	182.18	6	6	17
	182.18	6	6	21
	182.18	6	6	22
	182.18	6	6	23
	182.18	6	6	24
	182.18	6	6	25
	182.18	6	6	26
	182.18	6	6	27
Methotrexate	454.45	13	7	2
	454.45	13	7	3
	454.45	13	7	4
	454.45	13	7	4
	454.45	13	7	5
	454.45	13	7	7
	454.45	13	7	8
	454.45	13	7	9
	454.45	13	7	10
	454.45	13	7	11
	454.45	13	7	12
	454.45	13	7	13
	454.45	13	7	14
	454.45	13	7	15
	454.45	13	7	16
	454.45	13	7	17
	454.45	13	7	19
	454.45	13	7	21
	454.45	13	7	22
	454.45	13	7	23
	454.45	13	7	24
	454.45	13	7	25
	454.45	13	7	26

Methotrexate (continued)	454.45	13	7	27
Metoprolol-tartrate	267.37	4	2	3
	267.37	4	2	4
	267.37	4	2	5
	267.37	4	2	7
	267.37	4	2	8
	267.37	4	2	9
	267.37	4	2	11
	267.37	4	2	12
	267.37	4	2	13
	267.37	4	2	14
	267.37	4	2	15
	267.37	4	2	16
	267.37	4	2	17
	267.37	4	2	21
	267.37	4	2	22
	267.37	4	2	23
	267.37	4	2	24
	267.37	4	2	25
	267.37	4	2	27
Nadolol	309.41	5	4	3
	309.41	5	4	5
	309.41	5	4	7
	309.41	5	4	11
	309.41	5	4	13
	309.41	5	4	16
	309.41	5	4	21
	309.41	5	4	23
	309.41	5	4	24
	309.41	5	4	25
Naloxone	327.38	5	2	2
	327.38	5	2	3
	327.38	5	2	4
	327.38	5	2	5
	327.38	5	2	7
	327.38	5	2	8
	327.38	5	2	9
	327.38	5	2	10
	327.38	5	2	11
	327.38	5	2	12
	327.38	5	2	13
	327.38	5	2	14
	327.38	5	2	15
	327.38	5	2	16
	327.38	5	2	17
	327.38	5	2	19
	327.38	5	2	21
	327.38	5	2	22
	327.38	5	2	23
	327.38	5	2	24
	327.38	5	2	25
	327.38	5	2	27
Naproxen-sodium	230.27	3	1	2
	230.27	3	1	3
	230.27	3	1	4
	230.27	3	1	5
	230.27	3	1	7
	230.27	3	1	9
	230.27	3	1	11
	230.27	3	1	12
	230.27	3	1	13
	230.27	3	1	14
	230.27	3	1	16
	230.27	3	1	17

Naproxen-sodium (continued)	230.27	3	1	19
	230.27	3	1	21
	230.27	3	1	22
	230.27	3	1	23
	230.27	3	1	24
	230.27	3	1	25
	230.27	3	1	26
	230.27	3	1	27
Nortriptylene-HCl	263.39	1	1	3
	263.39	1	1	4
	263.39	1	1	5
	263.39	1	1	8
	263.39	1	1	9
	263.39	1	1	12
	263.39	1	1	13
	263.39	1	1	16
	263.39	1	1	17
	263.39	1	1	19
	263.39	1	1	21
	263.39	1	1	22
	263.39	1	1	24
	263.39	1	1	25
	263.39	1	1	27
Omeprazole	267.25	9	2	2
	267.25	9	2	3
	267.25	9	2	4
	267.25	9	2	5
	267.25	9	2	7
	267.25	9	2	8
	267.25	9	2	9
	267.25	9	2	11
	267.25	9	2	12
	267.25	9	2	13
	267.25	9	2	14
	267.25	9	2	15
	267.25	9	2	16
	267.25	9	2	17
	267.25	9	2	19
	267.25	9	2	20
	267.25	9	2	21
	267.25	9	2	22
	267.25	9	2	23
	267.25	9	2	24
	267.25	9	2	25
Phenytoin	451.49	10	2	3
	451.49	10	2	4
	451.49	10	2	5
	451.49	10	2	7
	451.49	10	2	8
	451.49	10	2	9
	451.49	10	2	10
	451.49	10	2	11
	451.49	10	2	12
	451.49	10	2	13
	451.49	10	2	14
	451.49	10	2	15
	451.49	10	2	16
	451.49	10	2	17
	451.49	10	2	19
	451.49	10	2	21
	451.49	10	2	22
	451.49	10	2	23
	451.49	10	2	24
	451.49	10	2	25

Phenytoin (continued)	451.49	10	2	26
	451.49	10	2	27
Piroxicam	331.35	7	2	3
	331.35	7	2	4
	331.35	7	2	5
	331.35	7	2	8
	331.35	7	2	9
	331.35	7	2	11
	331.35	7	2	12
	331.35	7	2	14
	331.35	7	2	15
	331.35	7	2	16
	331.35	7	2	17
	331.35	7	2	19
	331.35	7	2	21
	331.35	7	2	22
	331.35	7	2	23
	331.35	7	2	25
	331.35	7	2	26
	331.35	7	2	27
Potassium Bromide	119.00	1	0	16
Potassium Permangante	158.03	4	0	5
Prazosin	383.41	9	2	3
	383.41	9	2	4
	383.41	9	2	5
	383.41	9	2	7
	383.41	9	2	8
	383.41	9	2	9
	383.41	9	2	11
	383.41	9	2	12
	383.41	9	2	13
	383.41	9	2	14
	383.41	9	2	15
	383.41	9	2	16
	383.41	9	2	17
	383.41	9	2	21
	383.41	9	2	23
	383.41	9	2	25
Propranolol-HCl	259.35	3	2	3
	259.35	3	2	4
	259.35	3	2	5
	259.35	3	2	7
	259.35	3	2	8
	259.35	3	2	9
	259.35	3	2	11
	259.35	3	2	12
	259.35	3	2	13
	259.35	3	2	14
	259.35	3	2	15
	259.35	3	2	16
	259.35	3	2	17
	259.35	3	2	19
	259.35	3	2	21
	259.35	3	2	22
	259.35	3	2	23
	259.35	3	2	24
	259.35	3	2	25
	259.35	3	2	27
Quinidine	324.43	4	1	3
	324.43	4	1	4
	324.43	4	1	5
	324.43	4	1	8
	324.43	4	1	9
	324.43	4	1	10

Quinidine (continued)	324.43	4	1	11
	324.43	4	1	12
	324.43	4	1	13
	324.43	4	1	14
	324.43	4	1	16
	324.43	4	1	17
	324.43	4	1	19
	324.43	4	1	20
	324.43	4	1	21
	324.43	4	1	22
	324.43	4	1	23
	324.43	4	1	25
Ranitidine-HCl	314.41	7	2	2
	314.41	7	2	3
	314.41	7	2	3
	314.41	7	2	5
	314.41	7	2	7
	314.41	7	2	8
	314.41	7	2	9
	314.41	7	2	11
	314.41	7	2	12
	314.41	7	2	13
	314.41	7	2	14
	314.41	7	2	15
	314.41	7	2	16
	314.41	7	2	17
	314.41	7	2	19
	314.41	7	2	21
	314.41	7	2	22
	314.41	7	2	23
	314.41	7	2	24
	314.41	7	2	25
	314.41	7	2	26
	314.41	7	2	27
Silver Nitrate	169.87	3	0	11
Sodium Thiosulfate	158.11	4	0	9
	158.11	4	0	12
	158.11	4	0	16
	158.11	4	0	17
	158.11	4	0	19
	158.11	4	0	25
Tenidap	320.76	5	2	9
	320.76	5	2	11
	320.76	5	2	12
	320.76	5	2	14
	320.76	5	2	23
Terfenadine	471.69	3	2	3
	471.69	3	2	4
	471.69	3	2	5
	471.69	3	2	9
	471.69	3	2	11
	471.69	3	2	12
	471.69	3	2	14
	471.69	3	2	15
	471.69	3	2	16
	471.69	3	2	17
	471.69	3	2	19
	471.69	3	2	21
	471.69	3	2	22
	471.69	3	2	23
	471.69	3	2	25
Testosterone	288.43	2	1	1
	288.43	2	1	3
	288.43	2	1	4

Testosterone (continued)	288.43	2	1	5
	288.43	2	1	7
	288.43	2	1	9
	288.43	2	1	10
	288.43	2	1	11
	288.43	2	1	12
	288.43	2	1	13
	288.43	2	1	15
	288.43	2	1	16
	288.43	2	1	17
	288.43	2	1	19
	288.43	2	1	20
	288.43	2	1	21
	288.43	2	1	22
	288.43	2	1	23
	288.43	2	1	25
	288.43	2	1	26
	288.43	2	1	27
Trovafloxacin	416.36	7	3	2
	416.36	7	3	3
	416.36	7	3	5
	416.36	7	3	9
	416.36	7	3	12
	416.36	7	3	14
	416.36	7	3	16
	416.36	7	3	21
	416.36	7	3	25
Valproic-acid	144.22	2	1	1
	144.22	2	1	3
	144.22	2	1	4
	144.22	2	1	5
	144.22	2	1	7
	144.22	2	1	8
	144.22	2	1	9
	144.22	2	1	10
	144.22	2	1	11
	144.22	2	1	12
	144.22	2	1	13
	144.22	2	1	14
	144.22	2	1	15
	144.22	2	1	16
	144.22	2	1	17
	144.22	2	1	19
	144.22	2	1	21
	144.22	2	1	22
	144.22	2	1	23
	144.22	2	1	24
	144.22	2	1	25
	144.22	2	1	27
Vinblastine	811.00	13	3	2
	811.00	13	3	3
	811.00	13	3	4
	811.00	13	3	5
	811.00	13	3	7
	811.00	13	3	8
	811.00	13	3	9
	811.00	13	3	10
	811.00	13	3	11
	811.00	13	3	12
	811.00	13	3	13
	811.00	13	3	14
	811.00	13	3	15
	811.00	13	3	16
	811.00	13	3	17

Vinblastine (continued)	811.00	13	3	19
	811.00	13	3	21
	811.00	13	3	22
	811.00	13	3	23
	811.00	13	3	24
	811.00	13	3	25
	811.00	13	3	26
Zinc Chloride	136.29	0	0	5
	136.29	0	0	13
	136.29	0	0	15
	136.29	0	0	21
	136.29	0	0	25
Ziprasidone	412.95	5	1	3
	412.95	5	1	4
	412.95	5	1	5
	412.95	5	1	7
	412.95	5	1	11
	412.95	5	1	13
	412.95	5	1	14
	412.95	5	1	16
	412.95	5	1	19
	412.95	5	1	21
	412.95	5	1	22
	412.95	5	1	23
	412.95	5	1	25

Appendix D

Table 16: ANN Testing Data

Chemical Name	Molecular Weight (CC)	Hydrogen Acceptors (DD)	Hydrogen Donors (EE)	ANN Derived Disease (60/20/20)	ANN Rounded Disease (60/20/20)	ANN Derived Disease (70/15/15)	ANN Rounded Disease (70/15/15)	ANN Derived Disease (80/10/10)	ANN Rounded Disease (80/10/10)	Known Disease Category (II)
4-Ethyltoluene (622-96-8)	120.19	0	0	7.09	7	8.92	9	4.966	5	16
	120.19	0	0	2.35	5	11.89	12	6.62	7	22
	120.19	0	0	17.43	17	14.87	15	8.27	8	23
Benzanthrone (82-05-3)	230.26	1	0	10.56	11	4.5	5	8.11	8	5
	230.26	1	0	-2.21	-2	6	6	10.81	11	9
	230.26	1	0	-2.67	-3	7.5	8	13.52	14	12
	230.26	1	0	-2.64	-3	8	8	14.42	14	23
Salicylaldehyde (90-02-8)	122.12	2	1	6.37	6	8.5	9	6.4	6	7
	122.12	2	1	-5.76	-6	11.33	11	9.46	9	23
Benzyl Chloride (100-44-7)	156.58	0	0	10.56	11	4.5	5	8.12	8	8
	156.58	0	0	-2.21	-2	6	6	10.83	11	15
	156.58	0	0	-2.67	-3	7.5	8	13.54	14	22
n-Heptane (142-82-5)	100.20	0	0	3.35	3	4.5	5	5.36	5	3
	100.20	0	0	10.54	11	5.9	6	7.14	7	5
	100.20	0	0	11.27	11	7.3	7	8.93	9	16
	100.20	0	0	13.68	14	7.77	8	9.52	10	22
	100.20	0	0	15.14	15	8.24	8	10.12	10	23
n-Octane (Octane) (111-65-9)	114.22	0	0	4.19	4	2.24	2	7.32	7	16
	114.22	0	0	0.60	1	0.51	1	11.1	11	22
	114.22	0	0	3.36	3	-1.22	0	14.85	15	23
Propene (115-07-1)	42.07	0	0	5.88	6	5.66	6	4.85	5	3
	42.07	0	0	7.22	7	7.55	8	6.47	6	5
	42.07	0	0	3.98	4	9.44	9	8.09	8	8
	42.07	0	0	15.84	16	10.07	10	8.63	9	16
	42.07	0	0	13.37	13	10.7	11	9.17	9	19
	42.07	0	0	18.61	19	15.73	16	13.48	13	22
	42.07	0	0	-0.02	-0.02	1.26	1	1.08	1	23
Tetrahydrofuran (109-99-9)	72.10	1	0	7.09	7	8.92	9	4.98	5	3
	72.10	1	0	2.36	2	11.9	12	6.65	7	5
	72.10	1	0	17.43	17	14.87	15	8.31	8	8
	72.10	1	0	12.55	13	15.86	16	8.86	9	9
	72.10	1	0	8.55	9	16.86	17	9.42	9	12
	72.10	1	0	24.66	25	24.79	25	13.85	14	16
	72.10	1	0	0.64	1	1.98	2	1.1	1	22
	72.10	1	0	2.91	3	2.97	3	1.66	2	23
Triphenylene (217-59-4)	228.29	0	0	10.55	11	4.5	5	8.13	8	8
	228.29	0	0	-2.21	-2	6	6	10.84	11	23
Vinyl Acetate (108-05-4)	86.08	1	0	4.19	4	0.07	0	6.31	6	3
	86.08	1	0	0.60	1	0.09	0	8.42	8	5
	86.08	1	0	3.36	3	0.11	0	10.53	11	8
	86.08	1	0	2.27	2	0.11	0	11.23	11	9
	86.08	1	0	15.24	15	0.12	0	11.93	12	16
	86.08	1	0	20.61	21	0.17	0	17.55	18	19
	86.08	1	0	1.79	2	0.02	0	1.4	1	22
	86.08	1	0	-0.34	-0.3	0.03	0	2.1	2	23
	86.08	1	0	1.66	2	0.04	0	2.8	3	24

Bibliography

Abdel-Shafy, H. I., and Mansour, M. S. M. (2015). A review on polycyclic aromatic hydrocarbons: Source, environmental impact, effect on human health and remediaiton. *Egyptian Journal of Petroleum, 24(4),* 1-17. Retrieved from http://ac.els-cdn.com/S1110062114200237/1-s2.0-S1110062114200237-main.pdf?_tid=8951fa3c-d91e-11e5-9022-00000aacb35d&acdnat=1456116268_70f670a888bd72cdfb8508262ee74b47

Abraham, M. H., Grellier, P. L., Kamlet, M. J., Doherty R. M., Taft, R. W., and Abbound, J. M. (1989). The use of scales of hydrogen-bond acidity and basicity in organic chemistry. *Journal of the Portuguese Chemical Society, 31*(85), 85-92. Retrieved from http://www.spq.pt/magazines/rpq/318/article/1518/pdf

Adelman, D. E (2005). Article: The False Promise of the Genomics Revolution for Environmental Law. *The Harvard Environmental Law Review, 29*(1), 117-178. Retrieved from http://www.law.harvard.edu/students/orgs/elr/vol29_1/adelman.pdf

Adler, T., Shelton-Davenport, M., and Grossblatt, N. (2010). Use of Emerging Science and Technologies to Explore Epigenetic Mechanisms Underlying the Developmental Basis for Disease. *The Newsletter of the Standing Committee on Use of Emerging Science for Environmental Health Decisions, 1*. Retrieved from http://nas-sites.org/emergingscience/files/2011/07/epigenetics-newsletter-final1.pdf

Agatonovic-Kustrin, S., Beresford, R. (1999). Basic concepts of artificial neural network (ANN) modeling and its application in pharmaceutical. *Journal of Pharmaceutical and Biomedical Analysis, 22*(2000), 717-727. Retrieved from http://www.journals.elsevier.com/journal-of-pharmaceutical-and-biomedical-analysis

Albertson, G. (2006). Gene amplification in cancer. *Trends in Genetics, 22*(8), 447-445. Retrieved from http://www.sciencedirect.com/science/article/pii/S0168952506001752\

Alizadeh, B., Safdari, R., Zolnoori, M., & Bashiri, A. (2015). Developing an Intelligent System for Diagnosis of Asthma Based on Artificial Neural Networks. *Acta Informatica Medica, 23*(4), 220-223. Retrieved from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4584092/

American Society of Clinical Oncology (ASCO). (2012). Cancer.Net. *The Genetics of Cancer*. Retrieved from http://www.cancer.net/navigating-cancer-care/cancer-basics/genetics/genetics-cancer

American Chemical Council (ACC). (2011). ACC Prioritization Screening. Retrieved from http://www.americanchemistry.com/Prioritization-Document

American Chemical Society (ACS). (2016). Chemical Risk Assessment and Regulatory Decision Making. Retrieved from http://www.acs.org/content/dam/acsorg/policy/publicpolicies/promote/chemicalsmanagement/chemical-risk-assessment.pdf

Aurell, J., Gullett, B. K., and Yamamoto, D. P. (2012). Emissions from Open Burning of Simulated Military Waste from Forward Operating Bases. *Environmental Science and Technology*, *46*(20), 11004-11012. DOI:10.1021/es303131k

Baccarelli, A., Bollati, V. (2009). Epigenetics and environmental chemicals. *Current Opinion in Pediatrics*, *21*(2), 243-251. Retrieved from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3035853/

Barth, S. K., Dursa, E. K., Peterson, M. R., and Schneiderman, A. (2014). Prevalence of respiratory diseases among veterans of Operation Enduring Freedom and Operation Iraqi Freedom: results from the National Health Study for a New Generation of U.S. Veterans. *Military Medicine*, *179*(3), 241-245. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/24594456

Bonomini, F., Tengattini, S., Fabiano, A., Bianchi, R., and Rezzani, R. (2008). Atherosclerosis and oxidative stress. *Histol Histopathol*, 23(3), 381-90. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/18072094

Brown, J. A. (2015). Hazardous Chemicals and Occupational Diseases. HAZ-MAP Factsheet. Triphenylene. Retrieved from https://hazmap.nlm.nih.gov/category-details?table=copytblagents&id=7868

Brouch, E. (2014). Artificial neural Network Prediction of Chemical-Disease Relationships Using Readily Available Chemical Properties. Retrieved from AFIT Thesis AFIT-ENV-14-M-12.

Budinsky, R., Gollapudi, B., Albertini, R. J., Valentine, R., Stavanja, M., Teeguarden, J., Fensterheim, R., Rick, D., Lardie, T., McFadden, L., Green, A., and Recio, L. (2013). Nonlinear responses for chromosome and gene level effects induced by vinyl acetate monomer and its metabolite, acetaldehyde in TK6 cells. *Environmental and Molecular Mutagenesis*, *54*(9), 755-768. Retrieved from http://ctdbase.org/detail.go?type=reference&acc=24038327

Carcinogenic Risk In Occupational Settings (CRIOS). (2008). Genotoxicity: Sister Chromatid Exchange Test. Retrieved from http://www.crios.be/genotoxicitytests/sister_chromatid_exchange_test.htm

Center for Disease Control and Prevention (CDC). (2014). National Biomonitoring Program. Retrieved from http://www.cdc.gov/biomonitoring/about.html

Cheng, F., Sutariya, V. (2012). Applications of Artificial Neural Network Modeling in Drug Discovery. *Clinical and Experimental Pharmacology*, *2*(3), 1-2. Retrieved from http://dx.doi.org/10.4172/2161-1459.1000e113

Curseu, D., Surbu, D., Popa, M., Ionutas, A., and Czuczi, E. (2007). The Biologic Markers as useful tools in Epidemiological Research. 1st International Conference on Advancements of Medicine and Health Care through Technology. Retrieved from http://ie.utcluj.ro/files/acta/2007/Number4/Papers/P50308.pdf

Damborsky, J., Schultz, T. W. (1997). Comparison of the QSAR models for toxicity and biodegradability of anilines and phenols. *Chemosphere*, *34*(2), 429-446. Retrieved from http://www.sciencedirect.com/science/article/pii/S004565359600361X

Dangare, C. S., Apte, S. S. (2012). A data mining approach for prediction of heart disease using neural networks. *International journal of computer engineering & technology*, *3*(3), 30-40. Retrieved from http://www.iaeme.com/MasterAdmin/UploadFolder/A%20DATA%20MINING%20APPROACH%20FOR%20PREDICTION%20OF%20HEART%20DISEASE%20USING%20NEURAL%20NETWORKS/A%20DATA%20MINING%20APPROACH%20FOR%20PREDICTION%20OF%20HEART%20DISEASE%20USING%20NEURAL%20NETWORKS.pdf

Davis, A. P., Grodin, C. J., Lennon-Hopkins, K., Saraceni-Richards, C., Sciaky, D., King, B., Wiegers, T. C., and Mattingly, C. J. (2014). The Comparative Toxicogenomics Database's 10th year anniversary: Update 2015. *Nucleic Acids Research*, *43*(D1), 914-920. Retrieved from http://nar.oxfordjournals.org/content/43/D1/D914.long

Davis, A. P., Murphy, C. G., Saraceni-Richards, C., Rosenstein, M. C., Wiegers, T. C., and Mattingly, C. J. (2014). Comparative Toxicogenomics Database a knowledgebase and discovery tool for chemical–gene–disease networks. *Nucleic Acids Research*, *37*(1), 786-792. Retrieved from http://nar.oxfordjournals.org/content/37/suppl_1/D786.long

De Coster, S., and Van Larebeke, N. (2012). Endocrine-disrupting chemicals: associated disorders and mechanisms of action. *Journal Environment Public Health*, 2012. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/22991565

Defense Health Board (DHB) (2008). DHB Findings Pertaining to Health Risk Assessment, Burn Pit Exposures, Balad Air Base, Iraq. Retrieved from http://www.health.mil/~/media/MHS/Report%20Files/BaladHealthRiskAssessmentBoardFindingsRecommendation.ashx

DeStasio, E. (2015). The Ames Test. Retrieved from http://www.genetics-sa.org/education/pdf/GSA_DeStasio_Ames_Student_Resources.pdf

Drabløs, F., Feyzi, E., Aas, P. A., Vaagbø, C. B., Kavli, B., Bratlie, M. S., Peña-Diaz, J., Otterlei, M., Slupphaug, G., and Krokan, H. E. (2004). Alkylation damage in DNA and RNA—repair mechanisms and medical significance. *DNA Repair*, *3*(11), 1389-1407. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/15380096

Drent, M., Cobben, N. A. M., Henderson, R. F., Wouters, E. M. F., and Van Dieijen Visser, M., (1996). Usefulness of lactate dehydrogenase and its isoenzymes as indicators of lung damage and inflammation. *European Respiratory Journal*, *9*, 1736-1742. Retrieved from http://erj.ersjournals.com/content/erj/9/8/1736.full.pdf

Environmental Protection Agency (EPA). (2015). Summary of the Toxic Substances Control Act. Retrieved from http://www2.epa.gov/laws-regulations/summary-toxic-substances-control-act

Environmental Protection Agency (EPA). (2002). Emissions of Organic Air Toxics from Open Burning. Retrieved from http://nepis.epa.gov/Adobe/PDF/P1001G31.pdf

Environmental Protection Agency (EPA). (2000). Benzyl Chloride. Retrieved from http://www3.epa.gov/ttn/atw/hlthef/benzylch.html

Environmental Protection Agency (EPA). (2005). Endocrine Disruptor Screening Program (EDSP) Estrogen Receptor Bioactivity. Retrieved from http://www2.epa.gov/endocrine-disruption/endocrine-disruptor-screening-program-edsp-estrogen-receptor-bioactivity

Environmental Protection Agency (EPA). (1989). Health and Environmental Effects Document for n-Heptane. Retrieved from http://nepis.epa.gov/Exe/ZyNET.exe/900R0I00.txt?ZyActionD=ZyDocument&Client=EPA&Index=1986%20Thru%201990&Docs=&Query=%28carcinogen%29%20OR%20FNAME%3D%22900R0I00.txt%22%20AND%20FNAME%3D%22900R0I00.txt%22&Time=&EndTime=&SearchMethod=1&TocRestrict=n&Toc=&TocEntry=&QField=&QFieldYear=&QFieldMonth=&QFieldDay=&UseQField=&IntQFieldOp=0&ExtQFieldOp=0&XmlQuery=&File=D%3A%5CZYFILES%5CINDEX%20DATA%5C86THRU90%5CTXT%5C00000016%5C900R0I00.txt&User=ANONYMOUS&Password=anonymous&SortMethod=h%7C-&MaximumDocuments=1&FuzzyDegree=0&ImageQuality=r75g8/r75g8/x150y150g16/i425&Display=p%7Cf&DefSeekPage=x&SearchBack=ZyActionL&Back=ZyActionS&BackDesc=Results%20page&MaximumPages=1&ZyEntry=60&SeekPage=b

Environmental Protection Agency (EPA). (2012). Toxicological Review of Tetrahydrofuran. Retrieved from http://nepis.epa.gov/Exe/ZyNET.exe/P100FSUC.TXT?ZyActionD=ZyDocument&Client=EPA&Index=2011+Thru+2015&Docs=&Query=&Time=&EndTime=&SearchMethod=1&TocRestrict=n&Toc=&TocEntry=&QField=&QFieldYear=&QFieldMonth=&QFieldDay=&IntQFieldOp=0&ExtQFieldOp=0&XmlQuery=&File=D%3A%5Czyfiles%5CIndex%20Data%5C11thru15%5CTxt%5C00000006%5CP100FSUC.txt&User=ANONYMOUS&Password=anonymous&SortMethod=h%7C&MaximumDocuments=1&FuzzyDegree=0&ImageQuality=r75g8/r75g8/x150y150g16/i425&Display=p%7Cf&DefSeekPage=x&SearchBack=ZyActionL&Back=ZyActionS&BackDesc=Results%20page&MaximumPages=1&ZyEntry=1&SeekPage=x&ZyPURL

European Commission (EC). (2014). ESIS; IUCLID Dataset, alpha-Chlorotoluene (100-44-7). Retrieved from http://toxnet.nlm.nih.gov/cgi-bin/sis/search2/f?./temp/~VlOHGy:1

European Chemicals Agency (ECHA). (n.d.). Substance Information. Triphenylene. Web. Retrieved from http://echa.europa.eu/en/substance-information/-/substanceinfo/100.005.385

European Chemicals Agency (ECHA). (2008). Risk Assessment. Vinyl Acetate. Retrieved from http://echa.europa.eu/documents/10162/6434698/orats_final_rar_vinylacetate_en.pdf

European Chemicals Agency (ECHA). (2009). Proposal for Harmonised Classification and Labeling. Tetrahydrofuran. Retrieved from http://echa.europa.eu/documents/10162/13626/clh_axvrep_france_tetrahydrofuran en.pdf

European Chemicals Agency (ECHA). (2010). Committee for Risk Assessment (RAC). Tetrahydrofuran. Retrieved from http://echa.europa.eu/documents/10162/13579/annex1_background_document_thf_cas109-99-9_en.pdf

European Chemicals Agency (ECHA). (2010). Proposal for Harmonised Classification and Labeling. Vinyl Acetate. Retrieved from http://echa.europa.eu/documents/10162/13626/vinyl_acetate_en.pdf

Freda, C. E., and Stoppani, A. O. M. (1970). Kinetics of Bovine Liver Aldehyde Dehydrogenase. Effect of Coenzyme & Aldehyde Structure. *Enzymologia*, 38(4), 225. Retrieved from http://toxnet.nlm.nih.gov/cgi-bin/sis/search2/f?./temp/~gt5FkE:3

Food and Drug Administration (FDA). (2011). Interpretation of the Term "Chemical Action" in the Definition of Device Under Section 201(h) of the Federal Food, Drug, and Cosmetic Act. Regulatory Information. Retrieved from http://www.fda.gov/RegulatoryInformation/Guidances/ucm259059.htm

Gershenson, C. (2016). Artificial Neural Networks for Beginners. Cornell University library. Retrieved from http://arxiv.org/pdf/cs/0308031.pdf

Gini, G., Lorenzini, M., Benfenati, E., Grasso, P., and Bruschi, M. (1999). Predictive Carcinogenicity: A Model for Aromatic Compounds, with Nitrogen-Containing Substituents, Based on Molecular Descriptors Using an Artificial Neural Network. *Journal of Chem. Inf. Comput. Sci. 39*(158), 1076-1080. Retrieved from http://home.deib.polimi.it/gini/papers/JCICS99.pdf

Halliwell, B (2007). Oxidative stress and cancer: have we moved forward?. *Biochemical Journal, 401*(1), 1-11. Retrieved from http://www.biochemj.org/content/401/1/1

Harrill A. H., Ross, P. K., Gatti, D. M., Threadgill, D. W., and Rusyn, I. (2009). Population-Based Discovery of Toxicogenomics Biomarkers for Hepatotoxicity Using a Laboratory Strain Diversity Panel. *Toxicological Sciences, 110*(1), 235-243. Retrieved from http://www.rusynlab.com/publications/2009/Harrill%20et%20al%202009%20Tox Sci.pdf

Heinzel, A., Perco, P., Mayer, G., Oberbauer, R., Lukas, A., and Mayer, B. (2014). From molecular signatures to predictive biomarkers: modeling disease pathophysiology and drug mechanism of action. *Frontiers in Cell and Developmental Biology, 2*(37). Retrieved from http://journal.frontiersin.org/article/10.3389/fcell.2014.00037/abstract

Hernandez, L. G., Van Steeg, H., Luijten, M., and Van Benthem, J. (2009). Mechanisms of non- genotoxic carcinogens and importance of a weight of evidence approach. *Mutation Research/Reviews in Mutation Research, 682*(2-3), 94-109. Retrieved from http://www.sciencedirect.com/science/article/pii/S1383574209000428

Hou, L., Zhang, X., Wang, D., and Baccarelli, A. (2012). Environmental chemical exposures and human epigenetics. *International Journal Epidemiology, 41*(1), 79-105. Retrieved from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3304523/

Institute of Medicine (IOM). (2011). Long-term health consequences of exposure to burn pits in Iraq and Afghanistan. Washington, DC: The National Academies Press. Retrieved from http://iom.nationalacademies.org/Reports/2011/Long-Term-Health-Consequences-of-Exposure-to-Burn-Pits-in-Iraq-and-Afghanistan.aspx

International Labour Office. (1971). Benzanthrone. *Encyclopedia of Occupational Health And Safety*. (Vol. I and II, pp. 166). New York, New York: McGraw-Hill Book Co. Retrieved from http://toxnet.nlm.nih.gov/cgiin/sis/search2/f?./temp/~xh1j88:1

International Programme on Chemical Safety (INCHEM). (1993). Biomarkers and Risk Assessment: Concepts and Principles. Retrieved from http://www.inchem.org/documents/ehc/ehc/ehc155.htm#SectionNumber:1.3

Irigaray, P., Belpomme, D. (2009). Basic properties and molecular mechanisms of exogenous chemical carcinogens. *Carcinogenesis, 31*(2), 135-148. Retrieved from http://carcin.oxfordjournals.org/content/31/2/135.full.pdf

Jain, A. K., Mao, J., and Mohiuddin, K. M. (1996). Artificial Neural Networks: A Tutorial. *Computer, 3*(29), 31-44. Retrieved from http://www.cogsci.ucsd.edu/~ajyu/Teaching/Cogs202_sp12/Readings/jain_ann96.pdf

Jainik-Spiechowicz, E., and Wyszynska, K. (1998). Genotoxicity of Fabrasol and its component: 4-ethyltoluene. *Mutation Research, 417*(2-3), 95-100. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9733933

Lane, L. (2002). Better Living Through Toxicogenomics. *The Scientist*. Retrieved from http://www.the-scientist.com/?articles.view/articleNo/13888/title/Better-Living-Through-Toxicogenomics-/

Larder, B., Wang, D., Revell, A., Montaner, J., Harrigan, R., De Wolf, F., Lange, J., Wegner, S., Ruiz, L., Perez-Elias, M. J., Emery, S., Gatell, J., Monforte, A., Torti, C., Zazzi, M., and Lane, C. (2007). The development of artificial neural networks to predict virological response to combination HIV therapy. *Antiviral Therapy, 12*, 15-24. Retrieved from http://www.hivrdi.org/pdf/Antiviral-Therapy-12-15-24-2007.pdf

LeBaron, M. J., Rasoulpour, R. J., Klapacz, J., Ellis-Hutchings, R. G., Hollnagel, H. M., and Gollapude, B. B. (2010). Epigenetics and chemical safety assessment. *Mutation Research/Reviews in Mutation Research, 705*(2), 83-95. Retrieved from http://www.sciencedirect.com/science/article/pii/S1383574210000487

Lee, J., Lucas, C. (2014). Artificial Neural Networks. Topics in Cognitive Modeling (PowerPoint). Retrieved from http://www.inf.ed.ac.uk/teaching/courses/tcm/lectures/tcm03_ann.pdf

Lemiux, P. M., Lutes, C. C., and Santoianni, D. A. (2004). Emissions of Organic Air Toxics from Open Burning: A Comprehensive Review. *Progress in Energy and Combustion Science, 30*(1), 1-32. Retrieved from http://www.ingentaconnect.com/content/els/03601285/2004/00000030/00000001/art00061

Lipinski, C. A., Lombardo, F., Dominy, B. W., and Feeney, P. J. (2007). Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Advanced Drug Delivery Reviews, 46*(2001), 3-26. Retrieved from http://www.researchgate.net/publication/12072586_Lipinski_C._A._Lombardo_F._Dominy_B._W._Feeney_P._J._Experimental_and_Computational_Approaches_to_Estimate_Solubility_and_Permeability_in_Drug_Discovery_and_Development_Settings_Adv._Drug_Deliv._Rev._23_325_(1997)

Liu, L., Li, Y., and Tollefsbol, T. O. (2008). Gene-Environment Interactions and Epigenetic Basis of Human Disease. *Molecular Biology, 10*(1-2), 25-36. Retrieved from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2434999/

Llewellyn, L. E. (2007). Predictive toxinology: an initial foray using calculated molecular descriptors to describe toxicity using saxitoxins as a model. *Toxicon, 50*(7), 901-13. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/17675202

Kennedy, G., Spence, V. A., McLaren, M., Hill, A., Underwood, C., and Belch, J. J. F. (2005). Oxidative stress levels are raised in chronic fatigue syndrome and are associated with clinical symptoms. *Free Radical Biology and Medicine, 39*(5), 584-589. Retrieved from http://www.sciencedirect.com/science/article/pii/S0891584905002121

Khan, S., Mukhtar, H., and Pandya K. P. (1980). *n*-octane and *n*-nonane induced alterations in xenobiotic metabolising enzyme activities and lipid peroxidation of rat liver. *Toxicology, 16*(3), 239-245. Retrieved from http://www.sciencedirect.com/science/article/pii/0300483X80901201

Klaunig, J. E., Kamendulis, L. M., and Xu, Y. (2000). Epigenetic Mechanisms of Chemical Carcinogenesis. *Human & Experimental Toxicology, 19*(10), 543-555. Retrieved from http://www.belleonline.com/newsletters/volume9/vol9-1/n2v91.html

MAK Value Documentation. (2012). N-Heptane. *The MAK Collection for Occupational Health and Safety*, 166-176 Retrieved from http://onlinelibrary.wiley.com/doi/10.1002/3527600418.mb14282e0011/full#

Mattingly, C. J. (2008). Chemical Databases for Environmental Health and Clinical Research. *Toxicology Letters, 186*(1), 62-65. Retrieved from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2727625

Mattingly, C. J., Colby, G. T, Forrest, J. N., and Boyer, J. L. (2003). The Comparative Toxicogenomics Database (CTD). *Environmental Health Perspective, 111* (6), 793-795. Retrieved from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1241500/

Mattingly, C. J., Rosenstein, M. C., Colby, G. T, Forrest, J. N., and Boyer, J. L. (2006). The Comparative Toxicogenomics Database (CTD): A Resource for Comparative Toxicological Studies. *Journal of Experimental Zoology, 305A*(9), 689-692. Retrieved from http://onlinelibrary.wiley.com/doi/10.1002/jez.a.307/abstract;jsessionid=3810CC8EBDEF4327E42FDA032AD14624.f04t04

Mayeux, R. (2004). Biomarkers: Potential Uses and Limitations. *NeuroRx, 1*(2), 182-188. Retrieved from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC534923/

Mayo Medical Laboratories. (1999). Test ID. Alkaline Phosphatase, Serum. Retrieved from http://www.mayomedicallaboratories.com/test-catalog/Clinical+and+Interpretive/8340

McGrowder, D. A., Fraser, Y. P., Gordon, L., Crawford, T. V., and Rawlins, J. M. (2011). Serum creatine kinase and lactate dehyrodgenase activities in patients with thyroid disorders. *Niger Journal Clinical Practice, 14*(4), 454-9. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/22248949

McHale, C. M., Zhang, L., Hubbard, A. E., and Smith, M. T. (2010). Toxicogenomic profiling of chemically exposed humans in risk assessment. Mutation *Research/Review in Mutation Research*, *705*(3), 172-183. Retrieved from http://www.sciencedirect.com/science/article/pii/S1383574210000463

Moudgal, C. J., Garrahan, K., Brady-Roberts, E., Gavrelis, N., Arbogast, M., and Dun, S. (2008). Human health risk assessment database, "the NHSRC toxicity value database": supporting the risk assessment process at US EPA's National Homeland Security Research Center. *Toxicology and Applied Pharmacology*, *233*(1), 25-33. Retrieved from http://www.sciencedirect.com/science/article/pii/S0041008X08002780

National Center for Biotechnology Information (NCBI). (2016). *Benzanthrone*. *PubChem Compound Database*; CID=6697. Retrieved from http://pubchem.ncbi.nlm.nih.gov/compound/Benzanthrenone#section=Top

National Center for Biotechnology Information (NCBI). (2011). *Benzyl Chloride*. *PubChem Compound Database*; SID=131269097. Retrieved from http://pubchem.ncbi.nlm.nih.gov/substance/131269097#section=Top

National Center for Biotechnology Information (NCBI). (2016). *4-Ethytoluene*. *PubChem Compound Database*; CID=12160. Retrieved from http://pubchem.ncbi.nlm.nih.gov/compound/4-Ethyltoluene

National Center for Biotechnology Information (NCBI). (2016) *N-Heptane*. *PubChem Compound Database*; CID=8900. Retrieved from http://pubchem.ncbi.nlm.nih.gov/compound/8900

National Center for Biotechnology Information (NCBI). (2016). *Octane*. *PubChem Compound Database*; CID=356. Retrieved from http://pubchem.ncbi.nlm.nih.gov/compound/356#section=Top

National Center for Biotechnology Information (NCBI). (2016). Propene. *PubChem Compound Database*; CID=8252. Retrieved from http://pubchem.ncbi.nlm.nih.gov/compound/Propene#section=Top

National Center for Biotechnology Information (NCBI). (2016). Tetrahydrofuran. *PubChem Compound Database*; CID=8028. Retrieved from http://pubchem.ncbi.nlm.nih.gov/compound/8028

National Research Council. (1999). Toxicity of Military Smokes and Obscurants. Volume 3. Retrieved from http://www.nap.edu/catalog/9645/toxicity-of-military-smokes-and-obscurants-volume-3

National Toxicology Program (NTP). (2015). *Tox21*. Retrieved from http://ntp.niehs.nih.gov/results/hts/

New Jersey Department of Health (NJ DOH). (2007). *Hazardous Substance Fact Sheet: n-Heptane*. Retrieved from http://nj.gov/health/eoh/rtkweb/documents/fs/1339.pdf

New Jersey Department of Health (NJ DOH). (2004). *Hazardous Substance Fact Sheet: Propylene*. Retrieved from http://nj.gov/health/eoh/rtkweb/documents/fs/1609.pdf

Nendza, M., Wenzel, A. (2006). Discriminating Toxicant Classes by Mode of Action. *Environmental Science and Pollution Research, 13*(3), 192-203. Retrieved from https://www.researchgate.net/publication/7023875_Discriminating_Toxicant_Classes_by_Mode_of_Action_-_1_Ecotoxicity_Profiles

Organization for Economic Co-operation and Development (OECD). (2016). Introduction to (Quantitative) Structure Activity Relationships. Assessment of Chemicals. Retrieved from http://www.oecd.org/chemicalsafety/risk-assessment/introductiontoquantitativestructureactivityrelationships.htm

Ogu, C. C., Maxa, J. L. (2000). Drug interactions due to cytochrome P450. *Baylor University Med Cent, 13*(4), 421-423. Retrieved from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1312247/

Patel, V. P., Chu, C. T. (2011). Nuclear transport, oxidative stress, and neurodegeneration. *International Journal of Clinical Exp Pathology, 4*(3), 215-229. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3071655/

Pulliero, A., Cao, J., Vasques, L. D. R., and Pacchierotti, F. (2015). Genetic and Epigenetic Effects of Environmental Mutagens and Carcinogens. *BioMed Research International, 2015*(2015), 3 pages. Retrieved from http://www.hindawi.com/journals/bmri/2015/608054/

Rae, S. A., Wang, W. J., Wang, D., and Partridge, D. (1999). Artificial neural networks: a potential role in osteoporosis. *Journal of the Royal Society of Medicine, 92* (3), 119-122. Retrieved from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1297100/

Ramond, A., Godin-Ribuot, D., Ribuot, C., Totoson, P., Koritchneva, I., Cachot, S., Levy, P., and Joyeux-Faure, M. (2011). Oxidative stress mediates cardiac infraction aggravation induced by intermittent hypoxia. *Fundamental & Clinical Pharmacology, 27*(3), 252-261. Retrieved from http://onlinelibrary.wiley.com/doi/10.1111/j.1472-8206.2011.01015.x/abstract;jsessionid=9CA4E0F6EDDA7352CCC79110CDD319EA.f04t03

Repetto, M., Semprine, J., and Boveris, A. (2012). Lipid Peroxidation: Chemical Mechanism, Biological Implications and Analytical Determination. INTECH, 77(2), 3-30. Retrieved from http://cdn.intechopen.com/pdfs-wm/38477.pdf

Risen, J. (2010). Veterans Sound Alarm Over Burn-Pit Exposure. *The New York Times*, August 7, 2010, A10. Retrieved from http://www.nytimes.com/2010/08/07/us/07burn.html

Santos-Garcia, G., Varela, G., Novoa, N., & Jimenez, M. (2004). Prediction of postoperative morbidity after lung resection using an artificial neural network ensemble. *Artificial Intelligence in Medicine, 30*(1), 61-69. Retrieved from http://www.sciencedirect.com/science/article/pii/S0933365703000599

Scheringer, M., MacLeod, M., Matthies, M., Klasmeier, J., and Witt, J. (2006). Persistence criteria in the REACH Legislation: Critical Evaluation and Recommendations. Retrieved from http://www.reach-info.de/dokumente/gutachten_gesamtpersistenz.pdf

Sibanda, W., and Pretorius, P. (2012). Artificial Neural Networks- A Review of Applications of Neural Networks in the Modeling of HIV Epidemic. *International Journal of Computer Applications, 44*(16), 1-9. Retrieved from http://research.ijcaonline.org/volume44/number16/pxc3877438.pdf

Sidhu, S., Gullett, B., Striebich, R., Klosterman, J., Contreras, J., and DeVito, M. (2004). Endocrine disrupting chemical emissions for combustion source: diesel particulate emissions and domestic waste open burn emissions. *Atmospheric Environment, 39*(2005), 801-811. Retrieved from http://www.journals.elsevier.com/atmospheric-environment

Sikkema, J., De Bont, J. A. M., and Poolman, B. (1995). Mechanisms of Membrane Toxicity of Hydrocarbons. *Microbiological Reviews, 59*(2), 201-222. Retrieved from https://www.pubmedcentral.nih.gov/pmc/articles/PMC239360/pdf/590201.pdf

Silins, I., and Hogberg, J. (2011). Combined Toxic Exposures and Human Health: Biomarkers of Exposure and Effect. Toxicity of Hydrocarbons. *International Journal of Environmental Research and Public Health, 8*(3), 629-647. Retrieved from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3083662/

Singh, M. (2014). *Types of Biomarkers* [PowerPoint slides]. Retrieved from http://www.slideshare.net/samsinghpunjabi333/types-of-biomarkers-mandeep

Singh, N., Dhalla, A. K., Seneviratne, C., and Singal, P. K. (1995). Oxidative stress and Heart failure. *Molecular and Cellular Biochemistry, 147*(1), 77-81. Retrieved from http://link.springer.com/article/10.1007%2FBF00944786

Singh, R.P., Khanna, R., Kaw, J. L., and Das, M. (2003). Comparative effect of Benzanthrone and 3-bromobenzanthrone. *Molecular Toxicology, 77*(2), 94-90. Retrieved from http://link.springer.com/article/10.1007/s00204-002-0396-9

Shamisi, M. H., Assi, A. H., Hejase, and H., A., N. (2011). Using MATLAB to Develop Artificial Neural Network Models for Predicting Global Solar Radiation in Al Ain City – UAE. *Engineering Education and Research Using MATLAB,* ISBN: 978-953-307-656-0, InTech, Retrieved from http://www.intechopen.com/books/engineering-education-and-research-using-matlab/using-matlab-to-developartificial-neural-network-models-for-predicting-global-solar-radiation-in

Shatkin, J., Ranalli, B. (2007). Using Biomonitoring Data for Risk Characterization. Retrieved from http://www.cadmusgroup.com/wp-content/uploads/2013/02/Biomonitoring-Issue-Paper-FINAL_05_16_07.pdf

Shepherd, A. (1999). Protein Secondary Structure Prediction with Neural Networks: A Tutorial. *Computational Molecular Biology and Bioinformatics Discussion Group, Joint Resaerch School in Biomolecular Sciences, 21 March 1997.* Retrieved from http://www.biochem.ucl.ac.uk/~shepherd/sspred_tutorial/ss-index.html

Sheppard, D., McPhee, D., Darke, C., Shrethra, B., Moore, R., Jurewitz, A., and Gray, A. (1999). Predicting cytomegalovirus disease after renal transplantation: An artificial neural network approach. *International Journal of Medical Informatics, 54*(1), 55-76. Retrieved from http://www.sciencedirect.com/science/article/pii/S1386505698001695

Slupphaug, G., Kavli, B., and Krokan, H. E. (2003). The interacting pathways for prevention and repair of oxidative DNA damage. *Mutual Res, 531*(1-2), 231-51. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/14637258

Struck, S., Schmidt, U., Gruening, B., Jaeger, I. S., Hossbach, J., and Preissner, R. (2008). Toxicity versus potency: elucidation of toxicity properties discriminating between toxins, drugs, and natural compounds. *Genome Inform. 20*, 231-242. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/19425137

Swiercz, R., Rydzynski, K., Jajte, J., Stetliewicz, J., Majcherek, W. (2000). Studies on dermal, ocular, and respiratory effects of 4-ethyltoluene in experimental animals. *International Journal of Occupational and Environmental Health, 13*(4), 307-15. Retrieved from https://www.researchgate.net/publication/12056302_Studies_on_dermal_ocular_and_respiratory_effects_of_4-ethyltoluene_in_experimental_animals

Szutowski, M. M., and Rakoto, J. R. (2009). Multiplicity of n-Heptane oxidation pathways catalyzed by cytochrome P450. *Journal of Biochem Mol Toxicol. 23*(4), 287-294. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/19705363?dopt=Abstract

Tabb, M. M., Blumberg, B. (2013). New Modes of Action for Endocrine-Disrupting Chemicals. *Molecular Endocrinology, 20*(3). Retrieved from http://press.endocrine.org/doi/full/10.1210/me.2004-0513

Tennant, R. W. (2002). The National Center for Toxicogenomics: using new technologies to inform mechanistic toxicology. *Environmental Health Perspective, 110*(1), A8-10. Retrieved from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1240702/

Tewari, P., Roy, R., Mishra, S., Mandal, P., Yadav, A., Chaudhari, B. P., Chaturvedi, R. K., Dwivedi, P. D., Tripathi, A., and Das, M. (2015). Benzanthrone induced immunotoxicity via oxidative stress and inflammatory mediators in Balb/c mice. *Immunobiology, 220*(3), 369-81. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/25454808

Toxicology Data Network (TOXNET). (n.d.). 4-Ethyltoluene. *ChemIDplus a TOXNET database*. Web. Retrieved from http://chem.sis.nlm.nih.gov/chemidplus/rn/622-96-8

Toxicology Data Network (TOXNET). (2014). 4-Ethyltoluene. *TOXLINE*. Web. Retrieved from http://toxnet.nlm.nih.gov/cgi-bin/sis/search2/f?./temp/~gi4ugu:1

Toxicology Data Network (TOXNET). (2003). Benzanthrone. *Hazardous Substance Data Bank (HSDB)*. Web. Retrieved from http://toxnet.nlm.nih.gov/cgiin/sis/search2/r?dbs+hsdb:@term+@rn+@rel+82%2D05%2D3

Toxicology Data Network (TOXNET). (2015). Benzyl Chloride *ChemIDplus a TOXNET database*.. Web. Retrieved from http://chem.sis.nlm.nih.gov/chemidplus/rn/100-44-7

Toxicology Data Network (TOXNET). (2015). Benzyl Chloride. *Hazardous Substance Data Bank (HSDB)*. Web. Retrieved from http://toxnet.nlm.nih.gov/cgi-bin/sis/search2/r?dbs+hsdb:@term+@rn+@rel+100%2D44%2D7

Toxicology Data Network (TOXNET). (n.d.). Heptane. *ChemIDplus a TOXNET database*. Web. Retrieved from http://chem2.sis.nlm.nih.gov/chemidplus/rn/142-82-5

Toxicology Data Network (TOXNET). (2014). n-Heptane. *Hazardous Substance Data Bank (HSDB)*. Web. Retrieved from http://toxnet.nlm.nih.gov/cgi-bin/sis/search2/f?./temp/~bCwmQx:1

Toxicology Data Network (TOXNET). (n.d.). N-octane. *ChemIDplus a TOXNET database*. Web. Retrieved from http://chem2.sis.nlm.nih.gov/chemidplus/rn/111-65-9

Toxicology Data Network (TOXNET). (2014). n-Octane. *Hazardous Substance Data Bank (HSDB)*. Web. Retrieved from http://toxnet.nlm.nih.gov/cgi-bin/sis/search2/f?./temp/~AmkSQJ:1

Toxicology Data Network (TOXNET). (n.d.). Propylene. *ChemIDplus a TOXNET database*. Web. Retrieved from http://chem2.sis.nlm.nih.gov/chemidplus/rn/115-07-1

Toxicology Data Network (TOXNET). (1995). Propylene. *Chemical Carcinogenesis Research Information System (CCRIS)*. Web. Retrieved from http://toxnet.nlm.nih.gov/cgi-bin/sis/search2/r?dbs+ccris:@term+@rn++@rel+%22115%2D07%2D1%22

Toxicology Data Network (TOXNET). (2014). Propylene. *Hazardous Substance Data Bank (HSDB)*. Web. Retrieved from http://toxnet.nlm.nih.gov/cgi-bin/sis/search2/r?dbs+hsdb:@term+@rn+@rel+115%2D07%2D1

Toxicology Data Network (TOXNET). (2003). Salicylaldehyde. *Hazardous Substance Data Bank (HSDB)*. Web. Retrieved from http://toxnet.nlm.nih.gov/cgi-bin/sis/search2/r?dbs+hsdb:@term+@rn+@rel+90%2D02%2D8

Toxicology Data Network (TOXNET). (n.d.). Tetrahydrofuran. *ChemIDplus a TOXNET database*. Web. Retrieved from http://chem2.sis.nlm.nih.gov/chemidplus/rn/109-99-9

Toxicology Data Network (TOXNET). (2006). Tetrahydrofuran. *Chemical Carcinogenesis Research Information System (CCRIS)*. Web. Retrieved from http://toxnet.nlm.nih.gov/cgi-bin/sis/search2/r?dbs+ccris:@term+@rn++@rel+%22109%2D99%2D9%22

Toxicology Data Network (TOXNET). (2015). Tetrahydrofuran. *Hazardous Substance Data Bank (HSDB)*. Web. Retrieved from http://toxnet.nlm.nih.gov/cgi-bin/sis/search2/r?dbs+hsdb:@term+@rn+@rel+109%2D99%2D9

Toxicology Data Network (TOXNET). (n.d.). Triphenylene. *ChemIDplus a TOXNET database*. Web. Retrieved from http://chem2.sis.nlm.nih.gov/chemidplus/rn/217-59-4

Toxicology Data Network (TOXNET). (1985). Triphenylene. *Chemical Carcinogenesis Research Information System (CCRIS)*. Web. Retrieved from http://toxnet.nlm.nih.gov/cgi-bin/sis/search2/r?dbs+ccris:@term+@rn++@rel+%22217%2D59%2D4%22

Toxicology Data Network (TOXNET). (n.d.). Vinyl Acetate. *ChemIDplus a TOXNET database*. Web. Retrieved from http://chem2.sis.nlm.nih.gov/chemidplus/rn/108-05-4

Toxicology Data Network (TOXNET). (1998). Vinyl Acetate. *Genetic Toxicity Data Bank (GENE-TOX) a TOXNET database*. Web. Retrieved from http://toxnet.nlm.nih.gov/cgi-bin/sis/search2/r?dbs+genetox:@term+@rn+@rel+%22108%2D05%2D4%22

Toxicology Data Network (TOXNET). (2010). Vinyl Acetate. *Chemical Carcinogenesis Research Information System (CCRIS)*. Web. Retrieved from http://toxnet.nlm.nih.gov/cgi-bin/sis/search2/r?dbs+ccris:@term+@rn++@rel+%22108%2D05%2D4%22

Toxicology Data Network (TOXNET). (2011). Vinyl Acetate. *Hazardous Substance Data Bank (HSDB)*. Web. Retrieved from http://toxnet.nlm.nih.gov/cgi-bin/sis/search2/r?dbs+hsdb:@term+@rn+@rel+108%2D05%2D4

Tseng, W., Hung, L., Shieh, J., Abbod, M. F., and Lin, J. (2013). Hip fracture risk assessment: artificial neural network outperforms conditional logistic regression in an age- and sex-matched case control study. *BMC Musculoskelet Disord, 14*(207). Retrieved from http://www.sciencedirect.com/science/article/pii/S0933365703000599

U.S. National Library of Medicine. (2014). LiverTox. Alkylating Agents. Retrieved from http://livertox.nih.gov/AlkylatingAgents.htm

U.S. National Library of Medicine. (2014). Medline Plus. Lactate dehydrogenase test. Retrieved from https://www.nlm.nih.gov/medlineplus/ency/article/003471.htm

U.S. National Library of Medicine. (2015). Genetics Home Reference. Cystic Fibrosis. Retrieved from http://ghr.nlm.nih.gov/condition/cystic-fibrosis

Van Delft, J. H. M., Van Agen, E., Van Breda, S. G. J., Herwijnen, M. H., Staal, Y. C. M., and Kleinjans, J. C. S. (2004). Discrimination of genotoxic from non-genotoxic carcinogens by gene expression profiling. *Carcinogenesis, 25*(7), 1265-1276. Retrieved from http://carcin.oxfordjournals.org/content/25/7/1265.full.pdf

Van Leeuwen, D. M., Gottschalk, R. W. H., Schoeters, G., Van Larebeje, N. A., Nelen, V., Baeyens, W. F., Kleinjans, J. C. S., and Van Delft, J. H. M. (2008). Transcriptome Analysis in Peripheral Blood of Humans Exposed to Environmental Carcinogens: A Promising New Biomarker in Environmental Health Studies. *Environmental Health Prespect*, *116*(11). Retrieved from http://ehp.niehs.nih.gov/11401/

Veteran Affairs (VA). (2014). Public Health. Retrieved from http://www.publichealth.va.gov/PUBLICHEALTH/exposures/burnpits/index.asp

Villemin, D., Cherqaoui, D., and Mesbah, A. (1994). Predicting Carcinogenicity of Polycyclic Aromatic Hydrocarbons from Back Propagation Neural Network. *Journal of Chem. Inf. Comput. Sci, 34*, 1288-1293. Retrieved from https://www.researchgate.net/publication/220522566_Predicting_Carcinogenicity_of_Polycyclic_Aromatic_Hydrocarbons_from_Back-Propagation_Neural_Network

Vouk, B., Butler, C., Upton, A. C., Parke, D. V., and Casher, S. (1987). Study of structure-toxicity relationship by counterpropogation neural network. *Analytica Chimica Acta, 384* (1999), 319-332. Retrieved from http://dge.stanford.edu/SCOPE/SCOPE_30/SCOPE_30_1.5_Chapter5_99-114.pdf

Vracko, M., Novic, M., and Zupan, J. (1998). Use of Toxicity Data on Single Chemicals to Predict the Effects of Mixtures. *Methods for Assessing the Effects of Mixtures of Chemicals, 5*, 99-114.

Yorifuji, T., Noguchi, M., Tsuda, T., Suzuki, E., Takao, S., Kashima, S., and Yanagisawa, Y. (1987). Does open-air exposure to volatile organic compounds near a plastic recycling factory cause health effects?. *Journal of Occupational Health, 54*(2), 79-87. Retrieved from https://www.jstage.jst.go.jp/article/joh/54/2/54_11-0202-OA/_article

Walton, M. K. (2011). *Biomarkers and Qualification* [PowerPoint slides]. Retrieved from https://www2.rsna.org/re/QIBA_Annual_Meeting_2011/Index_files/WALTON.pdf

Weese, B. (2010). Issues Related to Burn Pits in Deployed Settings. Retrieved from http://www.cs.amedd.army.mil/AMEDDJournal/2010aprjun.pdf

Wiley-VCH. (1998). The MAK-Collection for Occupational Health and Safety. n-Heptane. Retrieved from http://onlinelibrary.wiley.com/doi/10.1002/3527600418.mb14282e0011/full

Woodall, B. (2012). "Emissions from Simulated Open Burning of Deployed US Military Waste". Retrieved from AFIT Thesis AFIT/GES/ENV/12-M05.

Woodall, B. D., Yamamoto, D. P., Gullett, B. K., and Touati, A. (2012). Emissions from Small-Scale Burns of Simulated Deployed U.S. Military Waste. *Environmental Science and Technology, 46*(20), 10997-11003. **DOI:** 10.1021/es3021556

World Health Organization (WHO). (2011). Biomarkers and Human Biomonitoring. Children's Health and the Environment. Retrieved from http://www.who.int/ceh/capacity/biomarkers.pdf

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 074-0188

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From – To)*
28-02-2016	Master's Thesis	May 2014 – March 2016

4. TITLE AND SUBTITLE
Using Artificial Neural Networks to Predict Disease Associations for Chemicals Present in Burn Pit Emissions

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Taylor, Amanda R., MSgt, USAF

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAMES(S) AND ADDRESS(S)
Air Force Institute of Technology
Graduate School of Engineering and Management (AFIT/ENV)
2950 Hobson Way, Building 640
WPAFB OH 45433-8865

8. PERFORMING ORGANIZATION REPORT NUMBER
AFIT-ENV-MS-16-M-189

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Intentionally Left Blank

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Distribution Statement A. Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES
This work is declared a work of the U.S. Government and is not subject to copyright protection in the United States.

14. ABSTRACT
In June of 2015, 27,378 of the 28,000 returning Operation Iraqi Freedom/Operation Enduring Freedom (OIF/OEF) veterans report being exposed to burn pits. According to Barth et al. (2014), 9,660 returning OIF/OEF veterans were diagnosed with respiratory diseases, to include asthma, bronchitis, and sinusitis, thus strengthening the need to develop decision support tools that can be used to understand the relationships between chemical exposure and disease. In this study an Artificial Neural Network (ANN) was used to predict the chemical-disease associations for burn pit constituents. Ten burn pit constituents were tested using varying hidden layers, similar chemical structure relationships, and three Training, Validation, and Testing (TVT) ratios. The ANN predicted misidentification rates of 73% or greater when the hidden layer size varied between 1 and 5. Misidentification rates of 75% or greater were observed for ANN simulations when the TVT ratios ranged from 60/20/20 to 80/10/10. ANN-based screening of chemical groups containing chemicals with benzene rings and chemicals containing hydrocarbon chains produced misidentification rates of 73% or greater, and R2 values of 0.0762 and lower. Hidden Layer size, TVT ratios, and chemical structure had little effect on the model's performance; additional training data is needed to improve the predictive capability of the ANN. The ANN-based screening of individual burn pit constituents produced several chemicals with R2 values greater than 0.8. These chemicals have been prioritized to further develop predictive ANN models for human health force support, resulting in the first research screening burn pit constituents with an ANN, and the first to prioritize burn pit emissions for future testing.

15. SUBJECT TERMS
ANN, Burn Pits, Disease Prediction, Health Associations

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Harper, Willie F., PhD, Civilian
U	U	U	UU	163	19b. TELEPHONE NUMBER *(Include area code)* (937) 255-6565, x3636 (willie.harper@afit.edu)

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. Z39-18